

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:28:33 ; Search time 62.2026 Seconds  
(without alignments)  
9357.426 Million cell updates/sec

Title: US-09-808-388-1

Perfect score: 20

Sequence: 1 caaaactaggtcaagggtca 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb\_ba:
- 2: gb\_hgt:
- 3: gb\_in:
- 4: gb\_om:
- 5: gb\_ov:
- 6: gb\_pat:
- 7: gb\_ph:
- 8: gb\_pr:
- 9: gb\_ro:
- 10: gb\_sts:
- 11: gb\_sv:
- 12: gb\_un:
- 13: gb\_vl:
- 14: gb\_vl:
- 15: em\_ba:
- 16: em\_fun:
- 17: em\_hum:
- 18: em\_in:
- 19: em\_mu:
- 20: em\_om:
- 21: em\_or:
- 22: em\_ov:
- 23: em\_pat:
- 24: em\_ph:
- 25: em\_pl:
- 26: em\_ro:
- 27: em\_sts:
- 28: em\_un:
- 29: em\_vi:
- 30: em\_hgt\_hum:
- 31: em\_hgt\_inv:
- 32: em\_hgt\_other:
- 33: em\_hgt\_mus:
- 34: em\_hgt\_pln:
- 35: em\_hgt\_rod:
- 36: em\_hgt\_man:
- 37: em\_hgt\_vrt:
- 38: em\_sv:
- 39: em\_hgtgo\_hum:
- 40: em\_hgtgo\_mus:
- 41: em\_hgtgo\_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	20	100.0	20	6	AX251573	Sequence
2	20	100.0	38	6	AX251574	Sequence
3	20	100.0	41	6	AX251575	Sequence
4	20	100.0	52	6	AX251576	Sequence
5	20	100.0	69	6	AX063386	Sequence
6	20	100.0	332	6	AX251578	Sequence
7	19	95.0	64	6	AX063387	Sequence
C 7	18.4	92.0	1634	10	AF169013	Rattus no
C 8	18.4	92.0	3587	10	BC026856	Mus muscu
C 9	18.4	92.0	5446	10	AF090697	Mus muscu
C 10	18.4	92.0	5546	10	AF090697	Mus muscu
C 11	18.4	92.0	5546	10	AF090696	Mus muscu
C 12	18.4	92.0	112269	2	AL845515	Mus muscu
C 13	18.4	92.0	204539	2	AC116994	Mus muscu
C 14	18.4	92.0	225433	2	AC125485	Mus muscu
C 15	18.4	92.0	245916	2	AL845492	Mus muscu
C 16	18.4	90.0	184850	9	AC017004	Homo sapi
C 17	18.4	90.0	208419	2	AC011993	Homo sapi
C 18	17.4	87.0	78132	2	AC021703	Homo sapi
C 19	17.4	87.0	100168	2	AC108974	Rattus no
C 20	17.4	87.0	114413	2	AC092286	Homo sapi
C 21	17.4	87.0	123386	8	F12F1	Arabidops
C 22	17.4	87.0	129108	9	AC011482	Homo sapi
C 23	17.4	87.0	153723	9	AC074101	Homo sapi
C 24	17.4	87.0	168080	2	AC009047	Homo sapi
C 25	17.4	87.0	180417	2	AC127766	Rattus no
C 26	17.4	87.0	181532	2	AC020700	Homo sapi
C 27	17.4	87.0	184213	9	AC073539	Homo sapi
C 28	17.4	87.0	198579	9	AC093536	Homo sapi
C 29	17.4	87.0	200242	2	AC102249	Mus muscu
C 30	17.4	87.0	200811	2	AC102559	Mus muscu
C 31	17.4	87.0	201879	2	AC122011	Mus muscu
C 32	17.4	87.0	214625	10	AC116582	Mus muscu
C 33	17.4	87.0	222179	2	AC119386	Rattus no
C 34	17.4	87.0	231225	2	AC123936	Mus muscu
C 35	17.4	87.0	233775	2	AC129336	Mus muscu
C 36	17.4	87.0	339044	2	AC122275	Mus muscu
C 37	17.4	85.0	25	6	AX472554	Sequence
C 38	17.4	85.0	94444	2	AC124875	Rattus no
C 39	17.4	85.0	113533	2	AC094170	Rattus no
C 40	17.4	85.0	124552	9	AC011403	Homo sapi
C 41	17.4	85.0	177057	2	AC095264	Rattus no
C 42	17.4	85.0	201681	2	AC119759	Rattus no
C 43	17.4	85.0	235092	2	AC099004	Rattus no
C 44	16.8	84.0	3710	10	MMU010949	Mus muscu
C 45	16.8	84.0	16963	8	SPAC16A10	Z97185 S.pombe chr

ALIGNMENTS

RESULT 1	AX251573	Sequence 1 from Patent WO0168845.	20 bp	DNA	linear	PAT-05-OCT-2001
LOCUS	AX251573					
DEFINITION	AX251573					
ACCESSION	AX251573					
VERSION	AX251573.1	GI:15984996				
KEYWORDS		synthetic construct.				
SOURCE		synthetic construct				
ORGANISM		artificial sequences.				
REFERENCE		1 (bases 1 to 20)				
AUTHORS		Masaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.				
TITLE		Inflammation-inducible hybrid promoters, vectors containing same				
JOURNAL		and uses thereof				
		Patent: WO 0168845-A 1 20-SEP-2001;				

```

FEATURES
  source
    Location/Qualifiers
      1..20
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="element PPPE"
BASE COUNT      9 a      4 c      4 g      3 t
ORIGIN
Query Match      100.0%; Score 20; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAGGTCA 20
Db 1 CAAACTAGGTCAGGTCA 20

RESULT 2
AX251574
LOCUS      AX251574      38 bp      DNA      linear      PAT 05-OCT-2001
DEFINITION Sequence 2 from Patent WO0168845.
ACCESSION  AX251574
VERSION     AX251574.1 GI:15984997
KEYWORDS
ORGANISM    synthetic construct.
            artificial sequences.
REFERENCE   1 (bases 1 to 38)
AUTHORS    Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
TITLE      Inflammation-inducible hybrid promoters, vectors containing same
            and uses thereof
JOURNAL
Patent: WO 0168845-A 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
  source
    Location/Qualifiers
      1..38
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="element PPPE"
BASE COUNT      17 a      7 c      8 g      6 t
ORIGIN
Query Match      100.0%; Score 20; DB 6; Length 38;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAGGTCA 20
Db 19 CAAACTAGGTCAGGTCA 38

RESULT 3
AX251575
LOCUS      AX251575      41 bp      DNA      linear      PAT 05-OCT-2001
DEFINITION Sequence 3 from Patent WO0168845.
ACCESSION  AX251575
VERSION     AX251575.1 GI:15984998
KEYWORDS
ORGANISM    synthetic construct.
            synthetic construct
            artificial sequences.
REFERENCE   1 (bases 1 to 41)
AUTHORS    Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
TITLE      Inflammation-inducible hybrid promoters, vectors containing same
            and uses thereof
JOURNAL
Patent: WO 0168845-A 30-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
  source
    Location/Qualifiers
      1..41
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="element PPPE"
BASE COUNT      18 a      8 c      8 g      7 t

ORIGIN
Query Match      100.0%; Score 20; DB 6; Length 41;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAGGTCA 20
Db 22 CAAACTAGGTCAGGTCA 41

RESULT 4
AX251576
LOCUS      AX251576      52 bp      DNA      linear      PAT 05-OCT-2001
DEFINITION Sequence 4 from Patent WO0168845.
ACCESSION  AX251576
VERSION     AX251576.1 GI:15984999
KEYWORDS
ORGANISM    synthetic construct.
            artificial sequences.
REFERENCE   1 (bases 1 to 52)
AUTHORS    Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
TITLE      Inflammation-inducible hybrid promoters, vectors containing same
            and uses thereof
JOURNAL
Patent: WO 0168845-A 4 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
  source
    Location/Qualifiers
      1..52
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="element PPPE"
BASE COUNT      19 a      11 c      11 g      11 t
ORIGIN
Query Match      100.0%; Score 20; DB 6; Length 52;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAGGTCA 20
Db 33 CAAACTAGGTCAGGTCA 52

RESULT 5
AX063386
LOCUS      AX063386      59 bp      DNA      linear      PAT 24-JAN-2001
DEFINITION Sequence 6 from Patent WO0078986.
ACCESSION  AX063386
VERSION     AX063386.1 GI:12541176
KEYWORDS
SOURCE      Human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 63)
AUTHORS    Parteil,R., Crouzet,J., Staelis,B. and Mahfoudi,A.
TITLE      Regulation system of expression using nuclear pparr receptors
            Patent: WO 0078986-A 6 28-DEC-2000;
            Aventis Pharma S.A. (FR)
FEATURES
  source
    Location/Qualifiers
      1..69
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
BASE COUNT      27 a      14 c      17 g      11 t
ORIGIN
Query Match      100.0%; Score 20; DB 6; Length 69;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAGGTCA 20

```



```

Db 17 CAAACTAGGTCAAAGGTCA 36

RESULT 6
AX251578
LOCUS
DEFINITION
Sequence 6 from Patent WO0168845.
ACCESSION
AX251578
VERSION
AX251578.1 GI:15985001
KEYWORDS
synthetic construct.
SOURCE
ORGANISM
artificial sequences.
REFERENCE
1 (bases 1 to 332)
AUTHORS
Masaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
TITLE
Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
JOURNAL
Patent: WO 0168845-A 6 20-SEP-2001;
Aventis Pharma S.A. (FR)
FEATURES
Location/Qualifiers
source
1..332
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoteur hybride PPRE/PLA2s"
BASE COUNT
96 a 91 c 82 g 63 t
ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 332;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAAGGTCA 20
|||||
Db 13 CAAACTAGGTCAAAGGTCA 32

RESULT 7
AX063387/c
LOCUS
DEFINITION
Sequence 7 from Patent WO0078986.
ACCESSION
AX063387
VERSION
AX063387.1 GI:12541177
KEYWORDS
human.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 64)
AUTHORS
Darteil,R., Crouzet,J., Staelens,B. and Mahfoudi,A.
TITLE
Regulation system of expression using nuclear ppar receptors
JOURNAL
Patent: WO 0078986-A 7 28-DEC-2000;
Aventis Pharma S.A. (FR)
FEATURES
Location/Qualifiers
source
1..64
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT
11 a 19 c 15 g 19 t
ORIGIN
Query Match 95.0%; Score 19; DB 6; Length 64;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAACTAGGTCAAAGGTCA 20
|||||
Db 49 AAAACTAGGTCAAAGGTCA 31

RESULT 8
AF169013/c
LOCUS
DEFINITION
Rattus norvegicus RNA binding protein NAPOR-3 mRNA, partial cds.
ACCESSION
AF169013

```

---

```

VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1634)
AUTHORS
Foleev,A.
TITLE
Direct Submission
JOURNAL
Submitted (12-JUL-1999) MPI for Neurobiology, Am Klopferspitze 18A,
Munich 82152, Germany
FEATURES
Location/Qualifiers
source
1..1634
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="brain"
/note="P5 brain library"
<1..682
CDS
/codon_start=2
/product="RNA binding protein NAPOR-3"
/protein_id="AA89096.1"
/db_xref="GI:9581852"
/translation="GLKGEDEKMGNTNANPLSTSSALGALTSFVAASPTNPSTAGAM
NSLTSGLTQLAGATVGLNNALVAQMLSGMAALNGGLGATGNTGTAGTMDALT
QAYSGIQYAAALPTLYSQSLQQQAAGSQKEGPEGANLFTYHLFPQFGQDILQM
FMPFGNIVSAKFIDKQTNLSKCFVSYDNPVSAQAALQAMNGFQIGMKRLKVLKR
SKNDSKEY"
BASE COUNT
462 a 351 c 342 g 479 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 1634;
Best Local Similarity 95.0%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAAGGTCA 20
|||||
Db 872 CAAACTAGGTCAAAGGTCA 853

RESULT 9
BC026856/c
LOCUS
DEFINITION
Mus musculus, CUG triplet repeat, RNA binding protein 2, clone
MGC:25225 IMAGE:4503295, mRNA, complete cds.
ACCESSION
BC026856
VERSION
BC026856.1 GI:20073072
KEYWORDS
MGC.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 3587)
AUTHORS
Strausberg,R.
TITLE
Direct Submission
JOURNAL
Submitted (04-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT
Contact: MGC help desk
Email: cgapbe@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 31 Row: 1 Column: 22  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6753779.

## FEATURES

Location/Qualifiers  
 1. .3587  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="MGC:25225 IMAGE:4503295"  
 /tissue\_type="Eye, retina, mouse strain C57Bl/6"  
 /clone\_lib="NIH\_MGC\_94"  
 /lab\_hosts="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 44. .1570

## CDS

/codon\_start=1  
 /product="CUG triplet repeat, RNA binding protein 2"  
 /protein\_id="AAH26856.1"  
 /db\_xref="GI:20073073"  
 /db\_xref="LocusID:14007"  
 /translation="MRCPKSAVTMRNEELLISNGTANKNGALDHSQDPDAIKMFV  
 QIPRSWSEKELPEFYGAVQINVLDRSQNPQSKGCCFVTYTRKAALQAQNA  
 LHNITLPGMHHPQMKPADSEKSNVDRKLFIGWSSKKNENDIRVWFSPFGQIEE  
 CRILRQGLSRGCAFTVFSTRMAQNAIKAMHQSOTMEGCSPIVVKFADTKDKEQ  
 RLQOQLAQOQQLNTATWNLTLGLGLTPQVLLALQOATSSNLGASGSGIQMAGMN  
 ALQCNATLAARAAATQATSTINAPLSSTSSALGALTSPVAASTNSTAGAMNS  
 LTSGLTQGLAGATVGLNNALAGMAALNGGLGTGLTNGTAGTMDALTQAYSIGIQ  
 YAAALPTLYSQSLQQAQSGKQEGEGANLFYHLRQEGDQDILQMWEPGNVI  
 SAKVFIDKQTNLSKCFGVSYDNFVSAQAIAQAMNGFQIGMKRLKVLQRKSNDSKPY  
 "

BASE COUNT 1026 a 798 c 742 g 1021 t  
 ORIGIN  
 Query Match 92.0%; Score 18.4; DB 10; Length 3587;  
 Best Local Similarity 95.0%; Pred. No. 52;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAAACTAGTCAAAAGTCA 20  
 |||||

Db 2862 CAAAACTAGTCAAAAGTCA 2843  
 |||||

RESULT 10  
 AF090697/c  
 LOCUS AF090697 5446 bp mRNA linear ROD 22-OCT-1999  
 DEFINITION Mus musculus apoptosis-related RNA binding protein (Napor-3) mRNA,  
 complete cds.  
 ACCESSION AF090697  
 VERSION AF090697.1 GI:4249673  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 5446)  
 AUTHORS Choi,D.K., Ito,T., Tsukahara,F., Hirai,M. and Sakaki,Y.  
 TITLE Developmentally-regulated expression of mNapor encoding an  
 apoptosis-induced ELAV-type RNA binding protein  
 GENE 237 (1), 135-142 (1999)  
 JOURNAL MEDLINE  
 PUBMED 99453769  
 10524244  
 REFERENCE 2 (bases 1 to 5446)  
 AUTHORS Choi,D.K., Ito,T. and Sakaki,Y.  
 TITLE Characterization and spatial distribution of the mNapor during  
 murine embryogenesis  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 5446)  
 AUTHORS Choi,D.K., Ito,T. and Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-SEP-1998) Human Genome Center, Institute of Medical  
 Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo  
 108, Japan

## FEATURES

Location/Qualifiers  
 1. .5446  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /map="A2-A3"  
 /tissue\_type="brain"  
 /dev\_stage="fetus"  
 1. .5446  
 /gene="Napor-3"  
 66. 1592  
 /gene="Napor-3"  
 /note="contains three copies of evolutionarily conserved  
 RNA recognition motif; developmentally regulated"  
 /codon\_start=1  
 /product="apoptosis-related RNA binding protein"  
 /protein\_id="AAD13764.1"  
 /db\_xref="GI:4249674"  
 /translation="MRCPKSAVTMRNEELLISNGTANKNGALDHSQDPDAIKMFV  
 QIPRSWSEKELPEFYGAVQINVLDRSQNPQSKGCCFVTYTRKAALQAQNA  
 LHNITLPGMHHPQMKPADSEKSNVDRKLFIGWSSKKNENDIRVWFSPFGQIEE  
 CRILRQGLSRGCAFTVFSTRMAQNAIKAMHQSOTMEGCSPIVVKFADTKDKEQ  
 RLQOQLAQOQQLNTATWNLTLGLGLTPQVLLALQOATSSNLGASGSGIQMAGMN  
 ALQCNATLAARAAATQATSTINAPLSSTSSALGALTSPVAASTNSTAGAMNS  
 LTSGLTQGLAGATVGLNNALAGMAALNGGLGTGLTNGTAGTMDALTQAYSIGIQ  
 YAAALPTLYSQSLQQAQSGKQEGEGANLFYHLRQEGDQDILQMWEPGNVI  
 SAKVFIDKQTNLSKCFGVSYDNFVSAQAIAQAMNGFQIGMKRLKVLQRKSNDSKPY  
 "

## gene

CDS  
 BASE COUNT 1511 a 1133 c 1094 g 1708 t  
 ORIGIN  
 Query Match 92.0%; Score 18.4; DB 10; Length 5446;  
 Best Local Similarity 95.0%; Pred. No. 51;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAAACTAGTCAAAAGTCA 20  
 |||||

Db 1780 CAAAACTAGTCAAAAGTCA 1761  
 |||||

RESULT 11  
 AF090696/c  
 LOCUS AF090696 5546 bp mRNA linear ROD 22-OCT-1999  
 DEFINITION Mus musculus apoptosis-related RNA binding protein (Napor-1) mRNA,  
 complete cds.  
 ACCESSION AF090696  
 VERSION AF090696.1 GI:4249671  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 5546)  
 AUTHORS Choi,D.K., Ito,T., Tsukahara,F., Hirai,M. and Sakaki,Y.  
 TITLE Developmentally-regulated expression of mNapor encoding an  
 apoptosis-induced ELAV-type RNA binding protein  
 GENE 237 (1), 135-142 (1999)  
 JOURNAL MEDLINE  
 PUBMED 99453769  
 10524244  
 REFERENCE 2 (bases 1 to 5546)  
 AUTHORS Choi,D.K., Ito,T. and Sakaki,Y.  
 TITLE Characterization and spatial distribution of the mNapor during  
 murine embryogenesis  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 5546)  
 AUTHORS Choi,D.K., Ito,T. and Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-SEP-1998) Human Genome Center, Institute of Medical  
 Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo  
 108, Japan

## FEATURES

Location/Qualifiers  
 1. .5546  
 /organism="Mus musculus"

```
/db_xref="taxon:10090"
/chrChromosome="2"
/map="A2-A3"
/tissue_type="brain"
/dev_stage="fetus"
1. .5546
/gene="Napor-1"
238..1692
/gene="Napor-1"
/notes="contains three copies of evolutionarily conserved
RNA recognition motif; developmentally regulated"
/codon_start=1
/product="apoptosis-related RNA binding protein"
/protein_id="A013763.1"
/db_xref="GI:4245672"
/translators="MNGALPHSDPDPAIKMVFQIPBSSEKELKEFEPYGVYQ
INVLDRSQNPSSKGCCTVFTTRAAALAAQNALNKTLPQMHPQMKPADSEK
NAVEDRLFIGVSKKCNENDIRVFPFGQIEECRLRGPDLGRGCAFTFTSTRAM
AQAIAKHQSQMTEGSSIVVVFADTKQKRRLOQLAQOQMLNTATWGLTG
LGLTPOYLLAQATSPVAASTPSTAGAAVNLSTSLGTLOGLAGATVGLNNIALA
NANPLSTGALGATGTNSTACTMDALTCAGVIAALPTLYSQSLQQOSRAGSO
GMAALNGLGATGLTNSTACTMDALTCAGVIAALPTLYSQSLQQOSRAGSO
KEGEGANLFIYHLPOBFGQDILQMPFGNVISAKVFDKQTNLSKCFGVSYDNP
VSAQAALQANNGFQIGMKRUKVQLKSKNDKPY"
BASE COUNT 1540 a 1176 c 1108 g 1722 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 5546;
Best Local Similarity 95.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
2Y 1 CAAACTAGGTCAAAGGTCA 20
3b 1880 CAAACTAGATCAAGGTCA 1861
RESULT 12
AL845515 112269 bp DNA linear HTG 16-AUG-2002
JOCUS Mus musculus chromosome 2 clone RP23-256D19, *** SEQUENCING IN
DEFINITION PROGRESS ***, 31 unordered pieces.
ACCESSION AL845515
VERSION AL845515.1 GI:22316245
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 112269)
Burton, J.
Direct Submission
Submitted (14-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM256D19
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 101632 bases at least Q40
Consensus quality: 101651 bases at least Q30
Consensus quality: 107942 bases at least Q20
Insert size: 109269; sum-of-contigs
Insert size: 186792; 1.2% error; agarose-fp
Quality coverage: 2.89x in Q20 bases; sum-of-contigs Quality
coverage: 2.22x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
```

\* consists of 31 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2026: contig of 2026 bp in length  
2027 2126: gap of 100 bp  
2127 4515: contig of 2389 bp in length  
4516 4615: gap of 100 bp  
4616 7195: contig of 2580 bp in length  
7196 7295: gap of 100 bp  
7296 14681: contig of 7386 bp in length  
14682 14781: gap of 100 bp  
14782 21050: contig of 6269 bp in length  
21051 21150: gap of 100 bp  
21151 23505: contig of 2355 bp in length  
23506 23605: gap of 100 bp  
23606 34391: contig of 10786 bp in length  
34392 34491: gap of 100 bp  
34492 37041: contig of 2550 bp in length  
37042 37141: gap of 100 bp  
37142 40346: contig of 3205 bp in length  
40347 40446: gap of 100 bp  
40447 42517: contig of 2071 bp in length  
42518 42617: gap of 100 bp  
42618 44873: contig of 2256 bp in length  
44874 44973: gap of 100 bp  
44974 50196: contig of 5223 bp in length  
50197 50296: gap of 100 bp  
50297 52455: contig of 2159 bp in length  
52456 52555: gap of 100 bp  
52556 54810: contig of 2255 bp in length  
54811 54910: gap of 100 bp  
54911 60983: contig of 6073 bp in length  
60984 61083: gap of 100 bp  
61084 63186: contig of 2103 bp in length  
63187 63286: gap of 100 bp  
63287 66133: contig of 2847 bp in length  
66134 66233: gap of 100 bp  
66234 68778: contig of 2545 bp in length  
68779 71439: contig of 2561 bp in length  
71440 71539: gap of 100 bp  
71540 74946: contig of 3407 bp in length  
74947 75046: gap of 100 bp  
75047 77413: contig of 2367 bp in length  
77414 77513: gap of 100 bp  
77514 79615: contig of 2102 bp in length  
79616 79715: gap of 100 bp  
79716 85282: contig of 5567 bp in length  
85283 85382: gap of 100 bp  
85383 92460: contig of 7078 bp in length  
92461 92560: gap of 100 bp  
92561 94871: contig of 2311 bp in length  
94872 94971: gap of 100 bp  
94972 98241: contig of 3270 bp in length  
98242 98341: gap of 100 bp  
98342 101349: contig of 3008 bp in length  
101350 101449: gap of 100 bp  
101450 104159: contig of 2710 bp in length  
104160 104259: gap of 100 bp  
104260 106320: contig of 2061 bp in length  
106321 106420: gap of 100 bp  
106421 109042: contig of 2622 bp in length  
109043 109142: gap of 100 bp  
109143 112269: contig of 3127 bp in length.

Location/Qualifiers  
1. .112269  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"

FEATURES  
Source

```

/misc_feature /clone="RP23-256D19"
1 /clone_lib="RPCI-23"
/note="assembly_fragment:00009"
2127.4515 /note="assembly_fragment:00017"
/misc_feature 4616.7195 /note="assembly_fragment:00080"
/misc_feature 7296.14681 /note="assembly_fragment:00136"
/misc_feature 14782.21050 /note="assembly_fragment:00161"
/misc_feature 21151.23505 /note="assembly_fragment:00194"
/misc_feature 23606.34391 /note="assembly_fragment:00228"
/misc_feature 34492.37041 /note="assembly_fragment:00259"
/misc_feature 37142.40346 /note="assembly_fragment:00299"
/misc_feature 40447.42517 /note="assembly_fragment:00315"
/misc_feature 42618.44873 /note="assembly_fragment:00317"
/misc_feature 44974.50195 /note="assembly_fragment:00365"
/misc_feature 50297.52455 /note="assembly_fragment:00379"
/misc_feature 52556.54810 /note="assembly_fragment:00418"
/misc_feature 54911.60983 /note="assembly_fragment:00492"
/misc_feature 61084.63186 /note="assembly_fragment:00529"
/misc_feature 63287.66133 /note="assembly_fragment:00533"
/misc_feature 66234.68778 /note="assembly_fragment:00599"
/misc_feature 68879.71439 /note="assembly_fragment:00642"
/misc_feature 71540.74946 /note="assembly_fragment:00690.0"
/misc_feature 75047.77413 /note="assembly_fragment:00702"
/misc_feature 77514.79615 /note="assembly_fragment:00733"
/misc_feature 79716.85282 /note="assembly_fragment:00764"
/misc_feature 85383.92460 /note="assembly_fragment:00784"
/misc_feature 92561.94871 /note="assembly_fragment:00791"
/misc_feature 94972.98241 /note="assembly_fragment:00821"
/misc_feature 98342.101349 /note="assembly_fragment:00825"
/misc_feature 101450.104159 /note="assembly_fragment:00830"
/misc_feature 104260.106320 /note="assembly_fragment:00838"
/misc_feature 106421.109042 /note="assembly_fragment:00862"
/misc_feature 109143.112269 /note="assembly_fragment:00886"
BASE COUNT 29854 a 23422 c 25097 g 30873 t 3023 others
ORIGIN

```

```

Query Match 92.0%; Score 18.4; DB 2; Length 112269;
Best Local Similarity 95.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 CAAAAGTCAAGGTCA 20
|||||

```

```

Db 58780 CAAAAGTCAAGGTCA 58799

RESULT 13
AC116994/c
LOCUS AC116994
DEFINITION HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
ACCESSION AC116994.1 GI:20043152
VERSION HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 204539)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus chromosome, clone Rp23-15703
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 204539)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferraira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Jones,C.,
Hagos,B., Horton,I., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehotzky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Navlor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Traversa,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L15762
Center clone name: L57_O3
----- Summary Statistics
Sequencing vector: Plasmid; N/A; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 194927 bases at least Q40
Consensus quality: 199654 bases at least Q30
Consensus quality: 200899 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 201539; sum-of-contigs
Quality coverage: 5.6 in Q20 bases; agarose-fp
Quality coverage: 5.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have

```

```

TITLE
JOURNAL
COMMENT

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

AC116994/c
LOCUS
DEFINITION

```

```

RESULT 13

```

```

Db 58780 CAAAAGTCAAGGTCA 58799

```

\* This sequence will be replaced.  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\*  
\* 24609 24708: gap of 100 bp  
\* 24709 25426: contig of 24608 bp in length  
\* 25427 25526: gap of 100 bp  
\* 25527 26055: contig of 718 bp in length  
\* 26056 26155: gap of 100 bp  
\* 26156 26772: contig of 529 bp in length  
\* 26773 26872: gap of 100 bp  
\* 26873 26970: contig of 617 bp in length  
\* 26971 27970: gap of 100 bp  
\* 27971 28070: contig of 1098 bp in length  
\* 28071 29324: gap of 100 bp  
\* 29325 29424: contig of 1254 bp in length  
\* 29425 30433: gap of 100 bp  
\* 30434 30533: contig of 1009 bp in length  
\* 30534 32405: gap of 100 bp  
\* 32406 32505: contig of 1872 bp in length  
\* 32506 33260: gap of 100 bp  
\* 33261 33360: contig of 755 bp in length  
\* 33361 35201: gap of 100 bp  
\* 35202 35301: contig of 1841 bp in length  
\* 35302 36304: gap of 100 bp  
\* 36305 36404: contig of 1003 bp in length  
\* 36405 37928: gap of 100 bp  
\* 37929 38028: contig of 1524 bp in length  
\* 38029 41081: gap of 100 bp  
\* 41082 41181: contig of 3053 bp in length  
\* 41182 42776: gap of 100 bp  
\* 42777 42876: contig of 1595 bp in length  
\* 42877 44443: gap of 100 bp  
\* 44444 44543: contig of 1567 bp in length  
\* 44544 46506: gap of 100 bp  
\* 46507 46606: contig of 1963 bp in length  
\* 46607 48831: gap of 100 bp  
\* 48832 48931: contig of 2225 bp in length  
\* 48932 51245: gap of 100 bp  
\* 51246 51345: contig of 2314 bp in length  
\* 51346 53740: gap of 100 bp  
\* 53741 53840: contig of 2395 bp in length  
\* 53841 59130: gap of 100 bp  
\* 59131 59230: contig of 2990 bp in length  
\* 59231 64768: gap of 100 bp  
\* 64769 64868: contig of 5538 bp in length  
\* 64869 72310: gap of 100 bp  
\* 72311 72410: contig of 7442 bp in length  
\* 72411 79109: gap of 100 bp  
\* 79110 79209: contig of 6699 bp in length  
\* 79210 93399: gap of 100 bp  
\* 93400 93499: contig of 14190 bp in length  
\* 93500 105638: gap of 100 bp  
\* 105639 105738: contig of 12133 bp in length  
\* 105739 118204: gap of 100 bp  
\* 118205 118304: contig of 12466 bp in length  
\* 118305 129803: gap of 100 bp  
\* 129804 129903: contig of 11499 bp in length  
\* 129904 147324: gap of 100 bp  
\* 147325 147424: contig of 17421 bp in length  
\* 147425 165479: gap of 100 bp  
\* 165480 165579: contig of 18055 bp in length  
\* 165580 182487: gap of 100 bp  
\* 182488 182587: contig of 16908 bp in length  
\* 182588 204539: gap of 100 bp  
\* 204540 219523: contig of 21952 bp in length.  
\*  
\* FEATURES  
\* Location/Qualifiers  
\* 1. 24609  
\* /organism="Mus musculus"  
\* /db\_xref="taxon:10090"  
\* /clone="RP23-15703"  
\* /clone\_lib="RP23-15703"  
\* 1. 24608  
\* /note="assembly\_fragment"

misc\_feature 1. 24608  
misc\_feature 1. 24609  
misc\_feature 1. 24708  
misc\_feature 1. 24709  
misc\_feature 1. 25426  
misc\_feature 1. 25427  
misc\_feature 1. 25526  
misc\_feature 1. 25527  
misc\_feature 1. 26055  
misc\_feature 1. 26056  
misc\_feature 1. 26155  
misc\_feature 1. 26156  
misc\_feature 1. 26772  
misc\_feature 1. 26773  
misc\_feature 1. 26872  
misc\_feature 1. 26873  
misc\_feature 1. 26970  
misc\_feature 1. 26971  
misc\_feature 1. 27970  
misc\_feature 1. 27971  
misc\_feature 1. 28070  
misc\_feature 1. 28071  
misc\_feature 1. 29324  
misc\_feature 1. 29325  
misc\_feature 1. 29424  
misc\_feature 1. 29425  
misc\_feature 1. 30433  
misc\_feature 1. 30434  
misc\_feature 1. 30533  
misc\_feature 1. 30534  
misc\_feature 1. 32405  
misc\_feature 1. 32406  
misc\_feature 1. 32505  
misc\_feature 1. 32506  
misc\_feature 1. 33260  
misc\_feature 1. 33261  
misc\_feature 1. 33360  
misc\_feature 1. 33361  
misc\_feature 1. 35201  
misc\_feature 1. 35202  
misc\_feature 1. 36304  
misc\_feature 1. 36305  
misc\_feature 1. 36404  
misc\_feature 1. 36405  
misc\_feature 1. 37928  
misc\_feature 1. 37929  
misc\_feature 1. 38028  
misc\_feature 1. 38029  
misc\_feature 1. 41081  
misc\_feature 1. 41082  
misc\_feature 1. 41181  
misc\_feature 1. 41182  
misc\_feature 1. 42776  
misc\_feature 1. 42777  
misc\_feature 1. 42876  
misc\_feature 1. 42877  
misc\_feature 1. 44443  
misc\_feature 1. 44444  
misc\_feature 1. 44543  
misc\_feature 1. 44544  
misc\_feature 1. 46506  
misc\_feature 1. 46507  
misc\_feature 1. 46606  
misc\_feature 1. 46607  
misc\_feature 1. 48831  
misc\_feature 1. 48832  
misc\_feature 1. 48931  
misc\_feature 1. 48932  
misc\_feature 1. 51245  
misc\_feature 1. 51246  
misc\_feature 1. 51345  
misc\_feature 1. 51346  
misc\_feature 1. 53740  
misc\_feature 1. 53741  
misc\_feature 1. 53840  
misc\_feature 1. 53841  
misc\_feature 1. 59130  
misc\_feature 1. 59131  
misc\_feature 1. 59230  
misc\_feature 1. 59231  
misc\_feature 1. 64768  
misc\_feature 1. 64769  
misc\_feature 1. 64868  
misc\_feature 1. 64869  
misc\_feature 1. 72310  
misc\_feature 1. 72311  
misc\_feature 1. 72410  
misc\_feature 1. 72411  
misc\_feature 1. 79109  
misc\_feature 1. 79110  
misc\_feature 1. 79209  
misc\_feature 1. 79210  
misc\_feature 1. 93399  
misc\_feature 1. 93400  
misc\_feature 1. 93499  
misc\_feature 1. 93500  
misc\_feature 1. 105638  
misc\_feature 1. 105639  
misc\_feature 1. 105738  
misc\_feature 1. 105739  
misc\_feature 1. 118204  
misc\_feature 1. 118205  
misc\_feature 1. 118304  
misc\_feature 1. 118305  
misc\_feature 1. 129803  
misc\_feature 1. 129804  
misc\_feature 1. 129903  
misc\_feature 1. 129904  
misc\_feature 1. 147324  
misc\_feature 1. 147325  
misc\_feature 1. 147424  
misc\_feature 1. 147425  
misc\_feature 1. 165479  
misc\_feature 1. 165480  
misc\_feature 1. 165579  
misc\_feature 1. 165580  
misc\_feature 1. 182487  
misc\_feature 1. 182488  
misc\_feature 1. 182587  
misc\_feature 1. 182588  
misc\_feature 1. 204539  
misc\_feature 1. 204540  
misc\_feature 1. 219523

clone\_end:SP6  
vector\_side:left  
24709..25426  
/note="assembly\_fragment"  
25527..26055  
/note="assembly\_fragment"  
26156..26772  
/note="assembly\_fragment"  
26873..27970  
/note="assembly\_fragment"  
28071..29324  
/note="assembly\_fragment"  
29425..30433  
/note="assembly\_fragment"  
30534..32405  
/note="assembly\_fragment"  
32506..33260  
/note="assembly\_fragment"  
33361..35201  
/note="assembly\_fragment"  
35302..36304  
/note="assembly\_fragment"  
36405..37928  
/note="assembly\_fragment"  
38029..41081  
/note="assembly\_fragment"  
41182..42776  
/note="assembly\_fragment"  
42877..44443  
/note="assembly\_fragment"  
44544..46506  
/note="assembly\_fragment"  
46607..48831  
/note="assembly\_fragment"  
48932..51245  
/note="assembly\_fragment"  
51346..53740  
/note="assembly\_fragment"  
53841..59130  
/note="assembly\_fragment"  
59231..64768  
/note="assembly\_fragment"  
64869..72310  
/note="assembly\_fragment"  
72411..79109  
/note="assembly\_fragment"  
79210..93399  
/note="assembly\_fragment"  
93500..105638  
/note="assembly\_fragment"  
105739..118204  
/note="assembly\_fragment"  
118305..129803  
/note="assembly\_fragment"  
129904..147324  
/note="assembly\_fragment"  
147425..165479  
/note="assembly\_fragment"  
165580..182487  
/note="assembly\_fragment"

Query Match 92.0%; Score 18.4; DB 2; length 204539;  
Best Local Similarity 95.0%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CAAACTAGGTCAGGTC 20  
Db 178629 CAAACTAGGTCAGGTC 178610  
RESULT 14  
AC125485/c AC125485 225433 bp DNA linear HTG 27-JUN-2002  
LOCUS AC125485 Mus musculus chromosome UNK clone RP23-16C13, WORKING DRAFT  
DEFINITION

```

SEQUENCE, 30 unordered pieces.
ACCESSION      AC125485
VERSION        AC125485.1  GI:21617766
KEYWORDS       HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE         house mouse.
ORGANISM       Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 225433)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 225433)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (27-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M BA0016C13
----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-terminator ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 214590 bases at least Q40
Consensus quality: 217676 bases at least Q30
Consensus quality: 219511 bases at least Q20
Insert size: 251000; agarose-ff
Insert size: 223743; sum-of-contigs
Quality coverage: 0.00 in Q20 bases; agarose-ff
Quality coverage: 6.35 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1511: contig of 1511 bp in length
* 1512 1611: gap of unknown length
* 1612 2923: contig of 1312 bp in length
* 2924 3023: gap of unknown length
* 3024 4722: contig of 1699 bp in length
* 4723 4822: gap of unknown length
* 4823 6381: contig of 1559 bp in length
* 6382 6482: gap of unknown length
* 6482 7910: contig of 1429 bp in length
* 7911 8010: gap of unknown length
* 8011 9503: contig of 1493 bp in length
* 9504 9604: gap of unknown length
* 9604 11058: contig of 1455 bp in length
* 11059 11158: gap of unknown length
* 11159 13254: contig of 2096 bp in length
* 13255 13354: gap of unknown length
* 13355 14882: contig of 1528 bp in length
* 14883 14983: gap of unknown length
* 14983 16254: contig of 1272 bp in length
* 16255 16354: gap of unknown length
* 16355 18120: contig of 1766 bp in length
* 18121 18220: gap of unknown length
* 18221 19610: contig of 1390 bp in length
* 19611 19710: gap of unknown length
* 19711 21191: contig of 1481 bp in length
*
SEQUENCE, 30 unordered pieces.
21192 21291: gap of unknown length
21292 23220: contig of 1929 bp in length
23221 23320: gap of unknown length
23321 26270: contig of 2950 bp in length
26271 26370: gap of unknown length
26371 28950: contig of 2580 bp in length
28951 29050: gap of unknown length
29051 32398: contig of 3248 bp in length
32399 36137: contig of 3739 bp in length
36138 36237: gap of unknown length
36238 39727: contig of 3490 bp in length
39728 39827: gap of unknown length
39828 44103: contig of 4276 bp in length
44104 44203: gap of unknown length
44204 47399: contig of 3196 bp in length
47400 47499: gap of unknown length
47500 54396: contig of 6897 bp in length
54397 60615: contig of 6119 bp in length
60616 60715: gap of unknown length
60716 67010: contig of 6295 bp in length
67011 67110: gap of unknown length
67111 76898: contig of 9788 bp in length
76899 84028: contig of 7030 bp in length
84029 84129: gap of unknown length
84130 97932: contig of 13804 bp in length
97933 98032: gap of unknown length
98033 130383: contig of 32351 bp in length
130384 130483: gap of unknown length
130484 171242: contig of 40759 bp in length
171243 171342: gap of unknown length
171343 225433: contig of 54091 bp in length.

FEATURES
Location/Qualifiers
1..225433
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-16C13"
misc_feature 1..1511
/note="assembly_name:Contig31"
misc_feature 1612..2923
/note="assembly_name:Contig33"
misc_feature 3024..4722
/note="assembly_name:Contig34"
misc_feature 4823..6381
/note="assembly_name:Contig36"
misc_feature 6482..7910
/note="assembly_name:Contig40"
misc_feature 8011..9503
/note="assembly_name:Contig42"
misc_feature 9604..11058
/note="assembly_name:Contig43"
misc_feature 11159..13254
/note="assembly_name:Contig44"
misc_feature 13355..14882
/note="assembly_name:Contig45"
misc_feature 14983..16254
/note="assembly_name:Contig46"
misc_feature 16355..18120
/note="assembly_name:Contig47"
misc_feature 18221..19610
/note="assembly_name:Contig48"
misc_feature 19711..21191
/note="assembly_name:Contig49"
misc_feature 21292..23220
/note="assembly_name:Contig50"
misc_feature 23321..26270
/note="assembly_name:Contig51"
misc_feature 26371..28950
/note="assembly_name:Contig52
clone_end:r7

```

```

vector_side:right"
misc_feature 29051..32298
/notes="assembly_name:Contig53"
misc_feature 32399..36137
/notes="assembly_name:Contig54"
misc_feature 36238..39727
/notes="assembly_name:Contig55"
misc_feature 39828..44103
/notes="assembly_name:Contig56"
misc_feature 44204..47399
/notes="assembly_name:Contig57"
misc_feature 47500..54396
/notes="assembly_name:Contig58"
misc_feature 54497..60615
/notes="assembly_name:Contig59"
misc_feature 60716..67010
/notes="assembly_name:Contig60"
misc_feature 67111..76898
/notes="assembly_name:Contig61"
misc_feature 76999..84028
/notes="assembly_name:Contig62"
misc_feature 84129..97932
/notes="assembly_name:Contig63"
misc_feature 98033..130383
/notes="assembly_name:Contig64"
misc_feature 130484..171242
/notes="assembly_name:Contig65"
misc_feature 171343..225433
/notes="assembly_name:Contig66"
BASE COUNT 66412 a 45098 c 45266 g 65747 t 2910 others
ORIGIN
Query Match 92.0%; Score 18.4; DB 2; Length 225433;
Best Local Similarity 95.0%; Pred. No. 37;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2y 1 CAAACTAGGTCAAAGGTCA 20
|||||
Db 160766 CAAACTAGATCAAGGTCA 160747

RESULT 15
AL845492/2 245916 bp DNA linear HTG 17-AUG-2002
LOCUS Mus musculus chromosome 2 clone RP23-222P7, *** SEQUENCING IN
DEFINITION PROGRESS ***, 11 unordered pieces.
ACCESSION AL845492
VERSION AL845492.2 GI:22416280
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 245916)
Plumb B.
Direct Submission
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 21, 2002 this sequence version replaced gi:22316222.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bM222P7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 243174 bases at least Q40
Consensus quality: 243936 bases at least Q30
Consensus quality: 244330 bases at least Q20

Insert size: 244916; sum-of-contigs
Insert size: 241095; 9.1% error; agarose-fp
Quality coverage: 6.99x in Q20 bases; sum-of-contigs Quality
coverage: 7.14x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 23764: contig of 23764 bp in length
* 23765 23864: gap of 100 bp
* 23865 54981: contig of 31117 bp in length
* 54982 55081: gap of 100 bp
* 55082 91767: contig of 36686 bp in length
* 91768 91867: gap of 100 bp
* 91868 94696: contig of 2829 bp in length
* 94697 94796: gap of 100 bp
* 94797 140303: contig of 45507 bp in length
* 140304 140403: gap of 100 bp
* 140404 144122: contig of 3719 bp in length
* 144123 144222: gap of 100 bp
* 144223 150853: contig of 6631 bp in length
* 150854 150953: gap of 100 bp
* 150954 192533: contig of 41580 bp in length
* 192534 192633: gap of 100 bp
* 192634 196792: contig of 4159 bp in length
* 196793 196892: gap of 100 bp
* 196893 200197: contig of 3305 bp in length
* 200198 200297: gap of 100 bp
* 200298 245916: contig of 45619 bp in length.
*
* Location/Qualifiers
* 1..245916
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /chromosomes="2"
* /clone="RP23-222P7"
* /clone_lib="RP23-23"
* 1..23764
* /notes="assembly fragment:02484"
* /fragment_chain:1"
misc_feature 23865..54981
/notes="assembly fragment:00096"
/fragment_chain:1"
misc_feature 55082..91767
/notes="assembly fragment:01352"
/fragment_chain:1"
misc_feature 91868..94696
/notes="assembly fragment:02266"
/fragment_chain:1"
misc_feature 94797..140303
/notes="assembly fragment:00090"
/fragment_chain:1"
misc_feature 140404..144122
/notes="assembly fragment:01481"
/fragment_chain:1"
misc_feature 144223..150853
/notes="assembly fragment:00446"
/fragment_chain:1"
misc_feature 150954..192533
/notes="assembly fragment:02940"
/fragment_chain:1"
misc_feature 192634..196792
/notes="assembly fragment:02221"
/fragment_chain:1"
misc_feature 196893..200197
/notes="assembly fragment:01308"
/fragment_chain:2"
misc_feature 200298..245916
/notes="assembly fragment:02802"

```

```
fragment chain:2
clone end:17
vector side:right"
BASE COUNT 71028 a 49815 c 49988 g 74099 t 1006 others
ORIGIN
Query Match 92.0%; Score 18.4; DB 2; Length 245916;
Best Local Similarity 95.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAAAACCTAGGTCAGGTC A 20
Db 181904 CAAAACCTAGGTCAGGTC A 181885
```

Search completed: February 19, 2003, 22:52:52  
Job time : 187.203 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:17:14 ; Search time 15.1708 Seconds  
(without alignments)  
671.453 Million cell updates/sec

Title: US-09-808-388-1

Perfect score: 20  
Sequence: 1 caaaactaggtcaagggtca 20

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Published Applications NA:

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	9	US-09-877-705A-67
2	20	100.0	20	9	US-09-877-705A-68
3	20	100.0	20	9	US-09-877-738A-67
4	20	100.0	20	9	US-09-877-738A-68
5	20	100.0	20	10	US-09-808-388-2
6	20	100.0	38	10	US-09-808-388-2
7	20	100.0	41	10	US-09-808-388-3
8	20	100.0	52	10	US-09-808-388-4
9	20	100.0	60	9	US-09-877-705A-142
10	20	100.0	332	10	US-09-877-738A-142
11	20	100.0	332	10	US-09-808-388-6
12	17	85.0	25	9	US-10-113-877-133
13	15.4	77.0	55795	10	US-09-880-107-1543
14	15.2	76.0	171	10	US-09-864-761-29451
15	15.2	76.0	272	10	US-09-394-093B-1029
16	15.2	76.0	573	10	US-09-864-761-12888
17	15.2	76.0	618	10	US-09-770-149-804
18	15.2	76.0	1021	10	US-09-881-752A-241
19	15.2	76.0	2000	9	US-09-938-842A-4842

C 20	15.2	76.0	3435	10	US-09-917-800A-1480	Sequence 1480, Ap
C 21	15.2	76.0	5129	9	US-09-938-842A-1803	Sequence 1803, Ap
C 22	15.2	76.0	14554	10	US-09-764-860-1054	Sequence 1054, Ap
C 23	15.2	76.0	14918	10	US-09-764-864-1766	Sequence 1766, Ap
C 24	15.2	76.0	29220	9	US-09-764-868-1312	Sequence 1312, Ap
C 25	15.2	76.0	29220	9	US-09-764-868-1313	Sequence 1313, Ap
C 26	15.2	76.0	1503841	9	US-09-946-807-1	Sequence 1, Appli
C 27	15.2	76.0	1503841	10	US-09-795-668-1	Sequence 1, Appli
C 28	15.2	76.0	1503841	10	US-09-795-668-1	Sequence 1, Appli
C 29	14.8	74.0	108	10	US-09-864-761-28906	Sequence 28906, A
C 30	14.8	74.0	260	10	US-09-878-574-5792	Sequence 5792, Ap
C 31	14.8	74.0	273	9	US-09-796-692-2542	Sequence 2542, Ap
C 32	14.8	74.0	350	10	US-09-728-446-1232	Sequence 1232, A
C 33	14.8	74.0	426	10	US-09-960-352-13817	Sequence 13817, A
C 34	14.8	74.0	448	10	US-09-864-761-2319	Sequence 2319, Ap
C 35	14.8	74.0	508	10	US-09-783-590-4401	Sequence 4401, Ap
C 36	14.8	74.0	509	10	US-09-864-761-12326	Sequence 12326, A
C 37	14.8	74.0	540	10	US-09-815-343-374	Sequence 374, App
C 38	14.8	74.0	565	10	US-09-864-761-6633	Sequence 6633, Ap
C 39	14.8	74.0	1431	10	US-09-822-849A-517	Sequence 517, App
C 40	14.8	74.0	2204	12	US-10-044-090-768	Sequence 768, App
C 41	14.8	74.0	2715	10	US-09-764-870-65	Sequence 65, Appli
C 42	14.8	74.0	3162	10	US-09-764-877-2937	Sequence 2937, Ap
C 43	14.8	74.0	14485	10	US-09-876-216-3	Sequence 3, Appli
C 44	14.8	74.0	50000	9	US-10-060-763-4	Sequence 4, Appli
C 45	14.8	74.0	50000	12	US-10-063-763-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-877-705A-67 EFD = 8/16/01

; Sequence 67, Application US/09877705A  
; Publication No. US20030008283A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, Jason  
; TITLE OF INVENTION: METHOD FOR SCREENING FOR DRUG CANDIDATES FOR MODULATING TRANSCRIPT  
; TITLE OF INVENTION: FACTOR ACTIVITY  
; FILE REFERENCE: 26757-704  
; CURRENT APPLICATION NUMBER: US/09/877,705A  
; CURRENT FILING DATE: 2001-08-16  
; NUMBER OF SEQ ID NOS: 162  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 67  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Transcription factor probe PP67  
US-09-877-705A-67

Query Match 100.0%; Score 20; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAAAACCTAGGTCAAAGGTCA 20  
Db 1 CAAAACCTAGGTCAAAGGTCA 20

RESULT 2  
US-09-877-705A-68/c  
; Sequence 68, Application US/09877705A  
; Publication No. US20030008283A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, Jason  
; TITLE OF INVENTION: METHOD FOR SCREENING FOR DRUG CANDIDATES FOR MODULATING TRANSCRIPT  
; TITLE OF INVENTION: FACTOR ACTIVITY  
; FILE REFERENCE: 26757-704  
; CURRENT APPLICATION NUMBER: US/09/877,705A  
; CURRENT FILING DATE: 2001-08-16  
; NUMBER OF SEQ ID NOS: 162

6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 68  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Transcription factor probe PP68  
US-09-877-705A-68

Query Match 100.0%; Score 20; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.23; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0

QY 1 CAAAACTAGGTCAAAGGTCA 20  
|||||  
DB 20 CAAAACTAGGTCAAAGGTCA 1

RESULT 3  
US-09-877-738A-67  
; Sequence 67, Application US/09877738A  
; Publication No. US20030022173A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, Jason  
; TITLE OF INVENTION: METHOD AND KIT FOR ISOLATING DNA PROBES THAT BIND TO ACTIVATED  
; TITLE OF INVENTION: TRANSCRIPTION FACTORS  
; FILE REFERENCE: 26757-701  
; CURRENT APPLICATION NUMBER: US/09/877,738A  
; PRIOR FILING DATE: 2001-06-01  
; NUMBER OF SEQ ID NOS: 162  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 67  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Transcription factor probe PP67  
US-09-877-738A-67

Query Match 100.0%; Score 20; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.23; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0

QY 1 CAAAACTAGGTCAAAGGTCA 20  
|||||  
DB 1 CAAAACTAGGTCAAAGGTCA 20

RESULT 4  
US-09-877-738A-68/c  
; Sequence 68, Application US/09877738A  
; Publication No. US20030022173A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, Jason  
; TITLE OF INVENTION: METHOD AND KIT FOR ISOLATING DNA PROBES THAT BIND TO ACTIVATED  
; TITLE OF INVENTION: TRANSCRIPTION FACTORS  
; FILE REFERENCE: 26757-701  
; CURRENT APPLICATION NUMBER: US/09/877,738A  
; PRIOR FILING DATE: 2001-06-01  
; NUMBER OF SEQ ID NOS: 162  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 68  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Transcription factor probe PP68  
US-09-877-738A-68

Query Match 100.0%; Score 20; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.23; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0

QY 1 CAAAACTAGGTCAAAGGTCA 20  
|||||  
DB 20 CAAAACTAGGTCAAAGGTCA 1

RESULT 5  
US-09-808-388-1  
; Sequence 1, Application US/09808388  
; Patent No. US20020081719A1  
; GENERAL INFORMATION:  
; APPLICANT: Massaad, Charbel  
; APPLICANT: Berenbaum, Francis  
; APPLICANT: Olivier, Jean-Luc  
; APPLICANT: Salvat, Colette  
; APPLICANT: Bereziat, Gilbert  
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
; TITLE OF INVENTION: their uses  
; FILE REFERENCE: ST00010  
; CURRENT APPLICATION NUMBER: US/09/808,388  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: FR/00/03262  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: US 60/196,959  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: PPRE element  
US-09-808-388-1

Query Match 100.0%; Score 20; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.23; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0

QY 1 CAAAACTAGGTCAAAGGTCA 20  
|||||  
DB 1 CAAAACTAGGTCAAAGGTCA 20

RESULT 6  
US-09-808-388-2  
; Sequence 2, Application US/09808388  
; Patent No. US20020081719A1  
; GENERAL INFORMATION:  
; APPLICANT: Massaad, Charbel  
; APPLICANT: Berenbaum, Francis  
; APPLICANT: Olivier, Jean-Luc  
; APPLICANT: Salvat, Colette  
; APPLICANT: Bereziat, Gilbert  
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
; TITLE OF INVENTION: their uses  
; FILE REFERENCE: ST00010  
; CURRENT APPLICATION NUMBER: US/09/808,388  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: FR/00/03262  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: US 60/196,959  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 38  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: PPRE element  
US-09-808-388-2

Query Match 100.0%; Score 20; DB 10; Length 38;

Best Local Similarity 100.0%; Pred. No. 0.26; Mismatches 0; Indels 0; Gaps 0;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAGGTC 20  
 DB 19 CAAACTAGGTCAGGTC 38

RESULT 7  
 US-09-808-388-3  
 ; Sequence 3, Application US/09808388  
 ; Patent No. US20020081719A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Massaad, Charbel  
 ; APPLICANT: Berenbaum, Francis  
 ; APPLICANT: Olivier, Jean-Luc  
 ; APPLICANT: Salvat, Colette  
 ; APPLICANT: Berzizat, Gilbert  
 ; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
 ; FILE REFERENCE: ST00010  
 ; CURRENT APPLICATION NUMBER: US/09/808,388  
 ; CURRENT FILING DATE: 2001-09-20  
 ; PRIOR APPLICATION NUMBER: FR/00/03262  
 ; PRIOR FILING DATE: 2000-03-14  
 ; PRIOR APPLICATION NUMBER: US 60/196,959  
 ; PRIOR FILING DATE: 2000-04-13  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 3  
 ; LENGTH: 41  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: PPRE element

US-09-808-388-3  
 Query Match 100.0%; Score 20; DB 10; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 0.27; Mismatches 0; Indels 0; Gaps 0;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAGGTC 20  
 DB 22 CAAACTAGGTCAGGTC 41

RESULT 8  
 US-09-808-388-4  
 ; Sequence 4, Application US/09808388  
 ; Patent No. US20020081719A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Massaad, Charbel  
 ; APPLICANT: Berenbaum, Francis  
 ; APPLICANT: Olivier, Jean-Luc  
 ; APPLICANT: Salvat, Colette  
 ; APPLICANT: Berzizat, Gilbert  
 ; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
 ; FILE REFERENCE: ST00010  
 ; CURRENT APPLICATION NUMBER: US/09/808,388  
 ; CURRENT FILING DATE: 2001-09-20  
 ; PRIOR APPLICATION NUMBER: FR/00/03262  
 ; PRIOR FILING DATE: 2000-03-14  
 ; PRIOR APPLICATION NUMBER: US 60/196,959  
 ; PRIOR FILING DATE: 2000-04-13  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 52  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: PPRE element

US-09-808-388-4  
 Query Match 100.0%; Score 20; DB 10; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 0.28; Mismatches 0; Indels 0; Gaps 0;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAGGTC 20  
 DB 33 CAAACTAGGTCAGGTC 52

RESULT 9  
 US-09-877-705A-142/c  
 ; Sequence 142, Application US/09877705A  
 ; Publication No. US20030008283A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Li, Jason  
 ; TITLE OF INVENTION: METHOD FOR SCREENING FOR DRUG CANDIDATES FOR MODULATING TRANSCRIPTION ACTIVITY  
 ; FILE REFERENCE: 26757-704  
 ; CURRENT APPLICATION NUMBER: US/09/877,705A  
 ; CURRENT FILING DATE: 2001-08-16  
 ; NUMBER OF SEQ ID NOS: 162  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 142  
 ; LENGTH: 60  
 ; TYPE: DNA  
 ; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Hybridization probe MP68

US-09-877-705A-142  
 Query Match 100.0%; Score 20; DB 9; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 0.28; Mismatches 0; Indels 0; Gaps 0;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAGGTC 20  
 DB 20 CAAACTAGGTCAGGTC 1

RESULT 10  
 US-09-877-738A-142/c  
 ; Sequence 142, Application US/09877738A  
 ; Publication No. US20030022173A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Li, Jason  
 ; TITLE OF INVENTION: METHOD AND KIT FOR ISOLATING DNA PROBES THAT BIND TO ACTIVATED TRANSCRIPTION FACTORS  
 ; FILE REFERENCE: 26757-701  
 ; CURRENT APPLICATION NUMBER: US/09/877,738A  
 ; CURRENT FILING DATE: 2001-06-01  
 ; NUMBER OF SEQ ID NOS: 162  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 142  
 ; LENGTH: 60  
 ; TYPE: DNA  
 ; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Hybridization probe MP68

US-09-877-738A-142  
 Query Match 100.0%; Score 20; DB 9; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 0.28; Mismatches 0; Indels 0; Gaps 0;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAGGTC 20  
 DB 20 CAAACTAGGTCAGGTC 1

RESULT 11  
 US-09-808-388-6

Sequence 6, Application US/09808388  
Patent No. US20020081719A1  
GENERAL INFORMATION:  
APPLICANT: Massaad, Charbel  
APPLICANT: Berenbaum, Francis  
APPLICANT: Olivier, Jean-Luc  
APPLICANT: Salvat, Colette  
APPLICANT: Berziat, Gilbert  
TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
TITLE OF INVENTION: their uses  
FILE REFERENCE: ST00010  
CURRENT APPLICATION NUMBER: US/09/808,388  
CURRENT FILING DATE: 2001-09-20  
PRIOR APPLICATION NUMBER: FR/00/03262  
PRIOR FILING DATE: 2000-03-14  
PRIOR APPLICATION NUMBER: US 60/196,959  
PRIOR FILING DATE: 2000-04-13  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 6  
LENGTH: 332  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: PPR/PLA2s hybrid promoter

US-09-808-388-6

Query Match 100.0%; Score 20; DB 10; Length 332;  
Best Local Similarity 100.0%; Pred. No. 0.38;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAGGTCA 20

Db 13 CAAACTAGGTCAAGGTCA 32

RESULT 12

US-10-113-877-133  
Sequence 133, Application US/10113877  
Patent No. US20020177218A1  
GENERAL INFORMATION:  
APPLICANT: Fang, Yu  
APPLICANT: Wang, Xiao-Yang  
APPLICANT: Turpin, Pierre  
TITLE OF INVENTION: Methods of detecting multiple DNA  
TITLE OF INVENTION: binding protein and DNA interactions in a sample, and  
TITLE OF INVENTION: devices, systems and kits for practicing the same.  
FILE REFERENCE: CLON-071  
CURRENT APPLICATION NUMBER: US/10/113,877  
CURRENT FILING DATE: 2002-03-29  
PRIOR APPLICATION NUMBER: 60/280,658  
PRIOR FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 60/314,330  
PRIOR FILING DATE: 2001-08-20  
NUMBER OF SEQ ID NOS: 192  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 133  
LENGTH: 25  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: oligonucleotide

US-10-113-877-133

Query Match 85.0%; Score 17; DB 9; Length 25;  
Best Local Similarity 100.0%; Pred. No. 8;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AACTAGGTCAAGGTCA 20

Db 3 AACTAGGTCAAGGTCA 19

RESULT 13

US-09-880-107-1543  
Sequence 1543, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gens Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1543  
LENGTH: 55795  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AF000573  
NAME/KEY: unsure  
LOCATION: (1)..(55795)  
OTHER INFORMATION: n = a or c or g or t

US-09-880-107-1543

Query Match 77.0%; Score 15.4; DB 10; Length 55795;  
Best Local Similarity 94.1%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AACTAGGTCAAGGTCA 20

Db 24241 AACTAGTCAAGGTCA 24257

RESULT 14

US-09-864-761-29451  
Sequence 29451, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30

DB 268 CAACACTAGGCCAAAGGCA 249

Search completed: February 20, 2003, 06:45:46  
Job time : 33.1708 secs

PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 29451  
LENGTH: 171  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC011159.3  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.52  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.48  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.64  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.5  
OTHER INFORMATION: NI HIT: AF121361.1, EVALUE 1.60e+00  
OTHER INFORMATION: SWISSPROT HIT: Q01955, EVALUE 2.30e+00  
OTHER INFORMATION: EST\_HUMAN HIT: N58147.1, EVALUE 2.30e+00  
JS-09-864-761-29451

Query Match 76.0%; Score 15.2; DB 10; Length 171;  
Best Local Similarity 85.0%; Pred No. 90;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2Y 1 CAAACTAGGTCAAAGGTCA 20  
|||||  
DB 151 CAAACCTATGTGAAAGGTCA 170

RESULT 15  
JS-09-294-093B-1029/c  
Sequence 1029, Application US/09294093B  
Patent No. US20010051335A1  
GENERAL INFORMATION:  
APPLICANT: Ito, Laura, Y.  
APPLICANT: Sherman, Bradley, K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL  
FILE REFERENCE: PL-0009 US  
CURRENT APPLICATION NUMBER: US/09/294,093B  
CURRENT FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: 60/082,567  
PRIOR FILING DATE: April 21, 1998  
NUMBER OF SEQ ID NOS: 6207  
SOFTWARE: PERL Program  
SEQ ID NO 1029  
LENGTH: 272  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. US20010051335A1 700343670H1  
JS-09-294-093B-1029

Query Match 76.0%; Score 15.2; DB 10; Length 272;  
Best Local Similarity 85.0%; Pred No. 98;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2Y 1 CAAACTAGGTCAAAGGTCA 20  
|||||

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:26:38 ; Search time 12.1555 Seconds  
(without alignments)  
3705.323 Million cell updates/sec

Title: US-09-808-388-1

Perfect score: 20

Sequence: 1 caaaactaggtcaagggtca 20

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SID32/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SID32/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SID32/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SID32/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SID32/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AAI64303	PPAR response elem
2	20	100.0	38	AAI64304	PPAR response elem
3	20	100.0	41	AAI64305	PPAR response elem
4	20	100.0	52	AAI64306	PPAR response elem
5	20	100.0	69	ABL58060	Human PPAR respons
6	20	100.0	332	AAI64308	Partial synthetic
c 7	19	95.0	64	ABL58061	Human PPAR respons
8	17	85.0	25	ABL98022	Cell-TRAP method a
c 9	16.8	84.0	506	AAI64307	Human reproductive

c 10	16.8	84.0	506	22	AAI62471	Human breast or ov
c 11	16.4	82.0	4232	21	AAF21792	Human breast and o
c 12	16.4	82.0	5644	24	ABX83490	Human cDNA differe
c 13	16	80.0	2698	23	ABL27652	Drosophila melanog
c 14	15.8	79.0	355	22	AAK59803	Human immune/haema
c 15	15.8	79.0	441	22	ABAI8904	Human nervous syst
c 16	15.8	79.0	516	21	AAAC09290	Partial rat alpha-
c 17	15.8	79.0	696	24	ABAI7263	Human ORFX polynuc
c 18	15.8	79.0	2012	22	AAK45153	cDNA encoding nove
c 19	15.8	79.0	2273	22	AAK94478	Human full-length
c 20	15.8	79.0	2622	22	AAI60347	Human polynucleoti
c 21	15.8	79.0	3011	23	AAK87623	DNA encoding novel
c 22	15.8	79.0	3276	22	AAK44965	cDNA encoding nove
c 23	15.8	79.0	6158	22	ABAI5973	Human nervous syst
c 24	15.8	79.0	6158	22	ABAI8902	Human nervous syst
c 25	15.8	79.0	6158	22	AAI05825	Human reproductive
c 26	15.8	79.0	6158	23	ABL98389	Human testicular a
c 27	15.4	77.0	386	24	ABK64462	Human benign prost
c 28	15.4	77.0	587	19	AAV56042	N. tabacum water c
c 29	15.4	77.0	587	19	AAV36214	cDNA sequence of t
c 30	15.4	77.0	55795	24	ABN95045	Gene #1543 used to
c 31	15.4	77.0	55795	24	ABL68242	Kidney cancer rela
c 32	15.4	77.0	55795	24	ABL68484	Kidney cancer rela
c 33	15.4	77.0	55795	24	ABL68863	Kidney cancer rela
c 34	15.2	76.0	59	22	AAH29316	Drosophila melanog
c 35	15.2	76.0	60	24	ABN46066	Human spliced tran
c 36	15.2	76.0	171	22	ABA74242	Human foetal liver
c 37	15.2	76.0	171	22	AAK22704	Human brain expres
c 38	15.2	76.0	171	22	AAK48871	Human bone marrow
c 39	15.2	76.0	272	24	ABU71855	Corn tassell-deriva
c 40	15.2	76.0	292	22	AAK40153	DNA encoding human
c 41	15.2	76.0	292	22	AAI00739	Human reproductive
c 42	15.2	76.0	321	23	ABV52055	Human prostate exp
c 43	15.2	76.0	340	22	AAK78102	Human immune/haema
c 44	15.2	76.0	340	22	AAK78104	Human immune/haema
c 45	15.2	76.0	434	22	AAK40505	DNA encoding human

ALIGNMENTS

RESULT 1

AAI64303  
ID AAI64303 standard; DNA; 20 BP.

XX AAI64303;

DT 15-NOV-2001 (first entry)

XX PPAR response element DRI.

XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;  
KW peroxisome proliferator activated receptor;  
KW secreted non-pancreatic phospholipase A2; ss.

OS Synthetic.

FN WO20016845-A2.

XX WO20016845-A2.

PD 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.

XX 14-MAR-2000; 2000FR-0003262.

PR 13-APR-2000; 2000US-0196959.

PA (AVET ) AVENTIS PHARMA SA.

PI Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;

XX WPI; 2001-582451/65.

XX

PT New hybrid promoter induced by inflammation, useful in gene therapy of  
 PT arthritis, comprises peroxisome proliferator activated receptor  
 PT response element and promoter of secreted phospholipase A2 -

PS Claim 3; Page 28; 52pp; French.

XX  
 CC The present invention relates to a hybrid promoter comprising (i) a PPAR  
 CC (peroxisome proliferator activated receptor) response element (PPRE); and  
 CC (ii) at least part of the promoter of the PLA2sIIA (secreted  
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to  
 CC regulate expression of therapeutic transgenes, for experimental,  
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for  
 CC treatment of arthritis, but also in bone, muscle, liver, heart, the  
 CC nervous system and tumours. The present sequence is a PPAR response  
 CC element, which was used to generate the hybrid promoter of the present  
 CC invention.

SQ Sequence 20 BP; 9 A; 4 C; 4 G; 3 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.2; Mismatches 0; Indels 0; Gaps 0;  
 Matches 20; Conservative 0;

QY 1 CAAACTAGGTCAAAGGTCA 20  
 |||||  
 Db 1 CAAACTAGGTCAAAGGTCA 20

## RESULT 2

AAI64304  
 ID AAI64304 standard; DNA; 38 BP.

XX  
 AC AAI64304;

DT 15-NOV-2001 (first entry)

DE PPAR response element (DR1)2 17.

XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
 KW cardiant; nontropic; promoter; arthritis; tumour; PLA2sIIA;  
 KW peroxisome proliferator activated receptor;  
 KW secreted non-pancreatic phospholipase A2; ss.

XX Synthetic.

XX WO200168845-A2.

XX 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.

XX 14-MAR-2000; 2000FR-0003262.

XX 13-APR-2000; 2000US-0196959.

XX (AVET ) AVENTIS PHARMA SA.

PI Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;

XX WPI; 2001-582451/65.

XX New hybrid promoter induced by inflammation, useful in gene therapy of  
 PT arthritis, comprises peroxisome proliferator activated receptor  
 PT response element and promoter of secreted phospholipase A2 -

XX Claim 4; Page 29; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR  
 CC (peroxisome proliferator activated receptor) response element (PPRE); and  
 CC (ii) at least part of the promoter of the PLA2sIIA (secreted  
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to  
 CC regulate expression of therapeutic transgenes, for experimental,  
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for  
 CC treatment of arthritis, but also in bone, muscle, liver, heart, the

CC nervous system and tumours. The present sequence is a PPAR response  
 CC element, which was used to generate the hybrid promoter of the present  
 CC invention.

SQ Sequence 38 BP; 17 A; 7 C; 8 G; 6 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 38;  
 Best Local Similarity 100.0%; Pred. No. 1.3; Mismatches 0; Indels 0; Gaps 0;  
 Matches 20; Conservative 0;

QY 1 CAAACTAGGTCAAAGGTCA 20  
 |||||  
 Db 19 CAAACTAGGTCAAAGGTCA 38

## RESULT 3

AAI64305  
 ID AAI64305 standard; DNA; 41 BP.

XX  
 AC AAI64305;

DT 15-NOV-2001 (first entry)

DE PPAR response element (DR1)2 21.

XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
 KW cardiant; nontropic; promoter; arthritis; tumour; PLA2sIIA;  
 KW peroxisome proliferator activated receptor;  
 KW secreted non-pancreatic phospholipase A2; ss.

XX Synthetic.

XX WO200168845-A2.

XX 20-SEP-2001.

XX 14-MAR-2001; 2001WO-FR00759.

XX 14-MAR-2000; 2000FR-0003262.

XX 13-APR-2000; 2000US-0196959.

XX (AVET ) AVENTIS PHARMA SA.

PI Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;

XX WPI; 2001-582451/65.

XX New hybrid promoter induced by inflammation, useful in gene therapy of  
 PT arthritis, comprises peroxisome proliferator activated receptor  
 PT response element and promoter of secreted phospholipase A2 -

XX Claim 4; Page 29; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR  
 CC (peroxisome proliferator activated receptor) response element (PPRE); and  
 CC (ii) at least part of the promoter of the PLA2sIIA (secreted  
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to  
 CC regulate expression of therapeutic transgenes, for experimental,  
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for  
 CC treatment of arthritis, but also in bone, muscle, liver, heart, the  
 CC nervous system and tumours. The present sequence is a PPAR response  
 CC element, which was used to generate the hybrid promoter of the present  
 CC invention.

XX Sequence 41 BP; 18 A; 8 C; 8 G; 7 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 1.3; Mismatches 0; Indels 0; Gaps 0;  
 Matches 20; Conservative 0;

QY 1 CAAACTAGGTCAAAGGTCA 20  
 |||||  
 Db 22 CAAACTAGGTCAAAGGTCA 41

KW peroxisome proliferator-activated receptor; promoter; PCR; primer; ss.  
 XX Homo sapiens.  
 XX WO2000078986-A1.  
 FN 28-DEC-2000.  
 XX 22-JUN-2000; 2000MO-FR01744.  
 XX 22-JUN-1999; 99ER-0007957.  
 PR 20-AUG-1999; 99US-0149721.  
 XX (AVET ) AVENTIS PHARMA SA.  
 XX Darteil R, Crouzet J, Staels B, Mahfoudi A;  
 PI WPI; 2001-091574/10.  
 DR  
 XX Composition providing inducible expression of a nucleic acid, useful in  
 PT gene therapy, uses minimal promoter with peroxisome  
 PT proliferator-activated receptor response elements -  
 XX  
 PS Example 1; Page 29; 94pp; French.  
 XX  
 CC The present invention relates to a composition (A) comprising a component  
 CC (A1) containing a nucleic acid (I) controlled by an inducible promoter  
 CC that consists of a PPAR (peroxisome proliferator-activated receptor)  
 CC response element (ABL58055) and a minimal promoter; and/or a component  
 CC (A2) comprising a nucleic acid encoding a PPAR under control of a  
 CC transcriptional promoter. (A), and vectors containing (A1) and (A2), are  
 CC used to express (I) in cells for expression of transgenic (I) for  
 CC experimental, clinical, therapeutic or diagnostic purposes. (I) encodes  
 CC an agriculturally useful, therapeutic, vaccinating or marker protein and  
 CC is most especially expressed in human muscle cells. Cells containing (A),  
 CC or the vectors, are used to identify PPAR ligands or to produce  
 CC transgenic animals for preclinical studies, analysis of bioavailability,  
 CC labelling etc. The present sequence is PCR primer, for human PPAR  
 CC response element consensus DR, which was used in an example from the  
 CC invention.  
 XX  
 XX Sequence 69 BP; 27 A; 14 C; 17 G; 11 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 20; DB 23; Length 69;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAAACCTAGGTCAAAGGTCA 20  
 |||||  
 DB 17 CAAAACCTAGGTCAAAGGTCA 36  
 RESULT 6  
 AAI64308  
 ID AAI64308 standard; DNA; 332 BP.  
 XX  
 XX AAI64308;  
 XX 15-NOV-2001 (first entry)  
 DT  
 XX  
 DE Partial synthetic PLA2sIIA gene promoter.  
 XX  
 KW PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
 KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;  
 KW peroxisome proliferator activated receptor;  
 XX secreted non-pancreatic phospholipase A2; ds.  
 XX  
 OS Synthetic.  
 XX  
 XX WO200168845-A2.  
 FN 20-SEP-2001.  
 PD  
 XX

RESULT 4  
 AAI64306  
 ID AAI64306 standard; DNA; 52 BP.  
 XX  
 XX AAI64306;  
 AC  
 XX 15-NOV-2001 (first entry)  
 DT  
 XX  
 DE PPAR response element (DR) 2 31.  
 XX  
 KW PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
 KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;  
 KW peroxisome proliferator activated receptor;  
 KW secreted non-pancreatic phospholipase A2; ss.  
 XX  
 OS Synthetic.  
 XX  
 XX WO200168845-A2.  
 FN 20-SEP-2001.  
 PD  
 XX  
 XX 14-MAR-2001; 2001WO-FR00759.  
 PF  
 XX 14-MAR-2000; 2000FR-0003262.  
 PR 13-APR-2000; 2000US-0196959.  
 XX  
 XX (AVET ) AVENTIS PHARMA SA.  
 XX  
 PI Masaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;  
 XX WPI; 2001-582451/65.  
 DR  
 XX New hybrid promoter induced by inflammation, useful in gene therapy of  
 PT arthritis, comprises peroxisome proliferator activated receptor  
 PT response element and promoter of secreted phospholipase A2 -  
 XX  
 XX Claim 4; Page 29; 52pp; French.  
 PS  
 XX  
 CC The present invention relates to a hybrid promoter comprising (i) a PPAR  
 CC (peroxisome proliferator activated receptor) response element (PPRE); and  
 CC (ii) at least part of the promoter of the PLA2sIIA (secreted  
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to  
 CC regulate expression of therapeutic transgenes, for experimental,  
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for  
 CC treatment of arthritis, but also in bone, muscle, liver, heart, the  
 CC nervous system and tumours. The present sequence is a PPAR response  
 CC element, which was used to generate the hybrid promoter of the present  
 CC invention.  
 XX  
 XX Sequence 52 BP; 19 A; 11 C; 11 G; 11 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 20; DB 22; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAAACCTAGGTCAAAGGTCA 20  
 |||||  
 DB 33 CAAAACCTAGGTCAAAGGTCA 52  
 RESULT 5  
 ABL58060  
 ID ABL58060 standard; DNA; 69 BP.  
 XX  
 XX ABL58060;  
 AC  
 XX 22-JUL-2002 (first entry)  
 DT  
 XX  
 DE Human PPAR response element consensus DR PCR primer 1RDA69.  
 XX  
 XX PPAR response element; PPAR; vaccine; gene therapy; human;  
 KW



PF 14-MAR-2001; 2001WO-FR00759.  
 XX  
 XX 14-MAR-2000; 2000FR-0003262.  
 PR 13-APR-2000; 2000US-0196959.  
 XX  
 XX (AVET ) AVENTIS PHARMA SA.  
 FA  
 XX Massaad C, Berenbaum F, Olivier J, Salvat C, Berezziat G;  
 PI WPI; 2001-582451/65.  
 XX  
 XX New hybrid promoter induced by inflammation, useful in gene therapy of  
 PT arthritis, comprises peroxisome proliferator activated receptor  
 PT response element and promoter of secreted phospholipase A2  
 XX  
 XX Disclosure; Page 51-52; 52pp; French.  
 PS  
 XX The present invention relates to a hybrid promoter comprising (i) a PPAR  
 CC (peroxisome proliferator activated receptor) response element (PPRE); and  
 CC (ii) at least part of the promoter of the PLA2SIIA (secreted  
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to  
 CC regulate expression of therapeutic transgenes, for experimental,  
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for  
 CC treatment of arthritis, but also in bone, muscle, liver, heart, the  
 CC nervous system and tumours. The present sequence is a partial synthetic  
 CC PLA2SIIA promoter sequence, which was used to generate the hybrid  
 CC promoter of the present invention.  
 XX  
 XX Sequence 332 BP; 96 A; 91 C; 82 G; 63 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 20; DB 22; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAACTAGGTCAAAGGTCA 20  
 DB 13 CAAACTAGGTCAAAGGTCA 32  
 RESULT 7  
 ABL58061/C  
 ID ABL58061 standard; DNA; 64 BP.  
 XX  
 AC ABL58061;  
 XX  
 XX 22-JUL-2002 (first entry)  
 DT  
 XX Human PPAR response element consensus DR PCR primer 2RDA64.  
 DE  
 XX PPAR response element; PPAR; vaccine; gene therapy; human;  
 KW peroxisome proliferator-activated receptor; promoter; PCR; primer; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200078986-A1.  
 PN  
 XX 28-DEC-2000.  
 PD  
 XX  
 XX 22-JUN-2000; 2000WO-FR01744.  
 PF  
 XX 22-JUN-1999; 92FR-0007957.  
 PR  
 XX 20-AUG-1999; 99US-0149721.  
 PT  
 XX (AVET ) AVENTIS PHARMA SA.  
 PA  
 XX Darteil R, Crouzet J, Staels B, Mahfoudi A;  
 PI WPI; 2001-091574/10.  
 DR  
 XX Composition providing inducible expression of a nucleic acid, useful in  
 PT gene therapy, uses minimal promoter with peroxisome  
 PT proliferator-activated receptor response elements -  
 XX

PS Example 1; Page 29; 94pp; French.  
 XX  
 XX The present invention relates to a composition (A) comprising a component  
 CC (A1) containing a nucleic acid (i) controlled by an inducible promoter  
 CC that consists of a PPAR (peroxisome proliferator-activated receptor)  
 CC response element (ABL58055) and a minimal promoter; and/or a component  
 CC (A2) comprising a nucleic acid encoding a PPAR under control of a  
 CC transcriptional promoter (A), and vectors containing (A1) and (A2), are  
 CC used to express (i) in cells for expression of transgenic (i) for  
 CC experimental, clinical, therapeutic or diagnostic purposes. (i) encodes  
 CC an agriculturally useful, therapeutic, vaccinating or marker protein and  
 CC is most especially expressed in human muscle cells. Cells containing (A),  
 CC or the vectors, are used to identify PPAR ligands or to produce  
 CC transgenic animals for preclinical studies, analysis of bioavailability,  
 CC labelling etc. The present sequence is PCR primer for human PPAR  
 CC response element consensus DR, which was used in an example from the  
 CC invention.  
 XX  
 SQ Sequence 64 BP; 11 A; 19 C; 15 G; 19 T; 0 other;  
 Query Match 95.0%; Score 19; DB 23; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 4.3;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 AAAACTAGGTCAAAGGTCA 20  
 DB 49 AAAACTAGGTCAAAGGTCA 31  
 RESULT 8  
 ABK98022  
 ID ABK98022 standard; DNA; 25 BP.  
 XX  
 AC ABK98022;  
 XX  
 XX 07-OCT-2002 (first entry)  
 DT  
 XX Cell-TRAP method associated Pit-1 oligonucleotide.  
 DE  
 XX Transcription factor; transcription factor-responsive element;  
 KW ds; TPRE; transcription activation; Cell-TRAP.  
 KW  
 XX Synthetic.  
 OS  
 XX WO200252039-A2.  
 PN  
 XX 04-JUL-2002.  
 PD  
 XX 21-DEC-2001; 2001WO-CA01861.  
 PF  
 XX 27-DEC-2000; 2000CA-2327581.  
 PR  
 XX (GENE-) GENEKA BIOTECHNOLOGY INC.  
 PA  
 XX Blais Y, Rousseau P, Leblanc B, Camato RN;  
 PI WPI; 2002-575388/61.  
 DR  
 XX A Cell-TRAP method, useful for producing or validating therapeutic  
 PT compounds, by employing a recombinant cell-based library that carry  
 PT constructs driven by a minimal promoter and a transcription  
 PT factor-responsive element -  
 XX  
 PS Disclosure; Page 25; 44pp; English.  
 XX  
 XX This invention relates to a cell-TRAP method for selecting and producing  
 CC a therapeutic compound which is presumed selective for, one or a  
 CC restricted set of given transcriptional pathways and cell types by  
 CC employing a recombinant cell-driven library that carries a construct  
 CC comprising a reporter gene driven by a minimal promoter and a  
 CC transcription factor-responsive element (TFRE). The invention also  
 CC comprises a method for validating a putative compound as a selective  
 CC therapeutic compound towards a transcription factor response element.  
 CC

CC The method of the invention is useful for determining the  
CC transcriptional activation pathways used by any compound that is  
CC biologically active in a cell. This method allows a global view of gene  
CC transcription activation in response to diverse stimuli in multiple  
CC environments and is a significant improvement over case-by-case  
CC approaches, which would be limited to certain aspects of gene  
CC activation. It permits to save on clinical trials by screening properly  
CC the compounds that would have a lesser probability of providing  
CC undesirable, even severe side effects. The present sequence  
CC represents a double stranded oligonucleotide probe recognised by a  
CC specific transcription factor which is used in the method of the  
CC invention.

XX Sequence 25 BP; 7 A; 7 C; 6 G; 5 T; 0 other;

Query Match 85.0%; Score:17; DB 24; Length 25;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 4 AACTAGGTCAAAGGTCA 20

Db 3 AACTAGGTCAAAGGTCA 19

RESULT 9

AA000071/c  
ID AAL00071 standard; cDNA; 506 BP.

AC AAL00071;

XX 21-NOV-2001 (first entry)

DE Human reproductive system related antigen cDNA SEQ ID NO: 72.

KW Human; reproductive system related antigen; reproductive system disorder;  
KW Cancer; gene therapy; ss.

XX Homo sapiens.

XX WO200155320-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.

```
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0359878.
XX XX
FA (HUMA-) HUMAN GENOME SCI INC.
XX XX
PI Rosen CA, Barash SC, Ruben SM;
XX XX
DR WPI; 2001-465570/50.
DR P-PSDB; AAM94101.
XX XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX XX
PS Claim 1; SEQ ID NO 72; 1297pp + Sequence Listing; English.
XX XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
CC invention.
XX XX
SQ Sequence 506 BP; 121 A; 99 C; 153 G; 128 T; 5 other;
Query Match 84.0%; Score 16.8; DB 22; Length 506;
Best Local Similarity 90.0%; Pred.No. 65;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 CAAAACTAGGTCAGAGGTCA 20
DB 413 CAAACACAGGTCAGATGCA 394
RESULT 10
AAI62471/c
ID AAI62471 standard; cDNA; 506 BP.
XX XX
AC AAI62471;
XX XX
DT 19-OCT-2001 (first entry)
XX XX
```

```
DE Human breast or ovarian antigen coding sequence SEQ ID NO: 15.
XX Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy;
KW ss.
XX Homo sapiens.
OS WC200155324-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01344.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 28-JUL-2000; 2000US-0223563.
PR 28-JUL-2000; 2000US-0225966.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
```

PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0244675.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.

PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI, 2001-488785/53.  
DR P-PSDB; AAM42244.  
XX  
PT New isolated nucleic acids and polypeptides, useful for diagnosing,  
PT treating and/or preventing human diseases and disorders -  
PS Claim 1; SEQ ID NO: 15; 520pp + Sequence Listing; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of ovarian and breast antigens. These are shown in  
CC AAI62467-AAI62572 and AAM42240-AAM42345. The sequences can be used in the  
CC diagnosis, prevention and treatment of breast and ovarian cancers, and  
CC their metastases. The present sequence is a coding sequence of the  
CC invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 506 BP; 121 A; 99 C; 153 G; 128 T; 5 other;  
Query Match 84.0%; Score 16.8; DB 22; Length 506;  
Best Local Similarity 90.0%; Pred. No. 65;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAAAAGTGGTCAAGGTCA 20  
DB 413 CAAAAGTGGTCAAGGTCA 394  
RESULT 11  
AAF21792/c  
ID AAF21792 standard; DNA; 4292 BP.  
XX  
AC AAF21792;  
XX  
DT 27-MAR-2001 (first entry)  
XX  
DE Human breast and ovarian cancer associated antigen gene SEQ ID 179.  
XX  
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;  
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;  
KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
KW cardiovascular disorder; wound healing; neurological disease; ds.  
XX Homo sapiens.  
XX  
XX WO200055173-A1.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000WO-US05881.  
XX  
XX 12-MAR-1999; 99US-0124270.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
XX WPI, 2000-611515/58.  
DR P-PSDB; AAB58889.  
XX

PT New human breast and ovarian cancer associated gene sequences and the  
PT polypeptides encoded by these genes, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -  
XX  
XX  
XX Claim 1; Page 614-615; 1299pp; English.  
XX  
XX Sequences AAF22614 - AAF22031 represent DNA sequences encoding human  
CC proteins AAF58711 - AAF59128. The DNA and protein sequences are  
CC associated with breast and ovarian cancer. Included in the invention are  
CC sequences AAF22032 - AAF22040 and AAF59129 which are used in the  
CC isolation and characterization of the DNA and protein sequences of the  
CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
CC or antagonist sequences exhibit cytostatic, immunosuppressive,  
CC neurotropic, neuroprotective, antiviral, antiallergic, hepatotropic;  
CC antidiabetic; antiinflammatory; antitumor; antileukemic; anticonvulsant;  
CC antibacterial; antifungal; antiparasitic and cardiant activity. The  
CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
CC particularly breast and ovarian cancer. The nucleic acid sequences,  
CC proteins, agonists and antagonists may also be used in the diagnosis,  
CC prevention and treatment of immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemias; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases.  
XX  
XX Sequence 4292 BP; 1286 A; 791 C; 748 G; 1465 T; 2 other;  
XX  
XX Query Match 82.0%; Score 16.4; DB 21; Length 4292;  
XX Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 3 AAAGTGTGCTCAAGGTCA 20  
XX |||||  
XX Db 465 AAAGTGTGCTCAAGGTCA 448  
XX  
XX RESULT 12  
XX ABK83490/c  
XX ID ABK83490 standard; cDNA; 5644 BP.  
XX  
XX AC ABK83490;  
XX  
XX DT 14-AUG-2002 (first entry)  
XX  
XX DE Human cDNA differentially expressed in granulocytic cells #61.  
XX  
XX KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
XX viral infection; parasitic infection; protozoal infection;  
XX fungal infection; sterile inflammatory disease; psoriasis;  
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
XX cardiac reperfusion injury; renal reperfusion injury; ARDS;  
XX adult respiratory distress syndrome; inflammatory bowel disease;  
XX Crohn's disease; ulcerative colitis; periodontal disease;  
XX granulocyte activation; chronic inflammation; allergy.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200228999-A2.  
XX  
XX PD 11-APR-2002.  
XX  
XX PF 03-OCT-2001; 2001WO-US30821.  
XX  
XX PR 03-OCT-2000; 2000US-237189P.  
XX  
XX PA (GENE-) GENE LOGIC INC.  
XX  
XX PI Bearer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX WPI; 2002-435328/46.  
XX  
XX DR  
XX

PT Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity -  
XX  
XX  
XX Claim 1; SEQ ID NO 61; 114pp; English.  
XX  
XX The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GCA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal  
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
CC periodontal disease; also bacterial infection, viral infection,  
CC parasitic infection, protozoal infection, fungal infection and M5 is  
CC useful for treating one of the above conditions. The present  
CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 5644 BP; 1573 A; 1206 C; 1129 G; 1736 T; 0 other;  
XX  
XX Query Match 82.0%; Score 16.4; DB 24; Length 5644;  
XX Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 3 AAAGTGTGCTCAAGGTCA 20  
XX |||||  
XX Db 1884 AAAGTGTGCTCAAGGTCA 1867  
XX  
XX RESULT 13  
XX ABL27652/c  
XX ID ABL27652 standard; DNA; 2698 BP.  
XX  
XX AC ABL27652;  
XX  
XX DT 26-MAR-2002 (first entry)  
XX  
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 34429.  
XX  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ds.  
XX  
XX OS Drosophila melanogaster.  
XX  
XX PN WO200171042-A2.  
XX  
XX PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX 11-JUL-2000; 2000US-0614150.  
PR  
XX (PEKE ) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW,  
PI  
XX WPI; 2001-656860/75.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 34429; 21pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (AB157737-AB172072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2698 BP; 731 A; 571 C; 548 G; 848 T; 0 other;  
SQ  
Query Match 80.0%; Score 16; DB 23; Length 2698;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 3 AACTAGGTCAAAGGT 18  
DB 1701 AACTAGGTCAAAGGT 1686  
RESULT 14  
AAK59803  
ID AAK59803 standard; cDNA; 355 BP.  
XX  
XX AAK59803;  
AC  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:4863.  
DE  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytosstatic; gene therapy; vaccine; metastasis; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157182-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX  
XX 04-FEB-2000; 2000US-0180628.  
PR  
XX 24-FEB-2000; 2000US-0184664.  
PR  
XX 02-MAR-2000; 2000US-0186350.  
PR  
XX 16-MAR-2000; 2000US-0189874.  
PR  
XX 17-MAR-2000; 2000US-0190076.  
PR  
XX 18-APR-2000; 2000US-0198123.  
PR  
XX 19-MAY-2000; 2000US-0205515.  
PR  
XX 07-JUN-2000; 2000US-0209467.  
PR  
XX 28-JUN-2000; 2000US-0214886.  
PR  
XX 30-JUN-2000; 2000US-0215135.  
PR

PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217486.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234598.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-02559678.  
PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-483426/52.

DR P-PSDB; HAM87022.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -

XX Claim 1; SEQ ID NO 4863; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in HAM82170 to AM91921. (I) have cytosolic  
CC activity, and can be used in gene therapy and vaccine production. (II)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.

XX Sequence 355 BP; 101 A; 64 C; 70 G; 114 T; 6 other;

Query Match 79.0%; Score 15.8; DB 22; Length 355;

Best Local Similarity 89.5%; Pred. No. 2e+02; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAGTAGGTCAGAGGCTCA 20

|||||

|||||

Db 296 AAAAAGTAGGTCAGAGGCTAA 314

RESULT 15

ABA18904

ID ABA18904 standard; DNA; 441 BP.

XX ABA18904;

XX 23-JAN-2002 (first entry)

XX Human nervous system related polynucleotide SEQ ID NO 11235.

Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
antiparkinsonian; antischizoid; antianaemic; antitubercular; cancer;  
antitumour; hepatotropic; cerebroprotective; antiinflammatory;  
antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

OS Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0217496.

XX 26-JUL-2000; 2000US-0220963.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226691.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234937.  
PR 25-SEP-2000; 2000US-0234938.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249266.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating nervous system  
XX cancers and metastases -  
XX  
XX Disclosure; SEQ ID NO 11235; 1701pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
XX (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy. The genes are  
XX isolated from a range of human tissues disclosed in the specification.  
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful  
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
XX and parasitic infections.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 441 BP; 82 A; 146 C; 123 G; 90 T; 0 Other;  
XX  
XX Query Match 79.0%; Score 15.8; DB 22; Length 441;



Best Local Similarity 89.5%; Pred. No. 2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY 2 AAAACTAGGTCAAAGGTCA 20  
| | | | | | | | | | | | | | | | | |  
Db 397 AAAACTGGTCCAGGTCA 415

Search completed: February 19, 2003, 21:22:34  
Job time : 13.155 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 20:28:04 ; Search time 2.55595 Seconds  
(without alignments)  
2399.710 Million cell updates/sec

Title: US-09-808-388-1

Perfect score: 20

Sequence: 1 caaaactagggtcaaggtca 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.4	77.0	587	4	US-09-053-702-3
C 2	15.2	76.0	3441	2	US-08-742-753-1
C 3	15.2	76.0	13865	3	US-09-009-217-11
C 4	15.2	76.0	13865	3	US-09-009-656-11
C 5	15.2	76.0	15894	1	US-08-348-891A-1
C 6	15.2	76.0	15894	1	US-08-905-817-1
C 7	15	75	2661	1	US-08-351-413-1
C 8	15	75	2661	2	US-09-025-583-1
C 9	15	75	4808	1	US-08-351-413-17
C 10	15	75	4808	2	US-09-025-583-17
C 11	14.8	74.0	11303	4	US-08-961-527-115
C 12	14.8	74.0	16595	4	US-09-146-053-7
C 13	14.2	71.0	865	4	US-09-328-111-128
C 14	14.2	71.0	1026	4	US-09-394-110A-3
C 15	14.2	71.0	1288	4	US-09-724-864-16
C 16	14.2	71.0	1846	4	US-09-336-536-37
C 17	14.2	71.0	2103	3	US-08-931-952-1
C 18	14.2	71.0	2103	3	US-08-272-247-1
C 19	14.2	71.0	2103	5	PCT-US95-08560-1
C 20	14.2	71.0	2849	4	US-09-221-017B-990
C 21	14.2	71.0	2964	2	US-08-846-790A-2
C 22	14.2	71.0	2964	3	US-08-935-333-2
C 23	14.2	71.0	3095	4	US-09-293-549-7
C 24	14.2	71.0	5521	4	US-08-975-762-48
C 25	14.2	71.0	5521	4	US-09-295-028-48
C 26	14.2	71.0	5521	4	US-09-106-582-48
C 27	14.2	71.0	5789	4	US-09-242-948-3
C 28	14.2	71.0	9707	4	US-08-961-527-164
C 29	14.2	71.0	45546	4	US-09-146-053-6
C 30	14.2	71.0	80246	4	US-09-078-294-4
C 31	14.2	71.0	80595	4	US-09-078-294-3
C 32	14.2	71.0	111282	4	US-09-754-250-3
C 33	14	70.0	1060	4	US-09-072-596-306
C 34	13.8	69.0	528	4	US-09-615-192A-137
C 35	13.8	69.0	545	2	US-08-975-316-74
C 36	13.8	69.0	545	4	US-09-615-192A-74
C 37	13.8	69.0	1072	4	US-09-280-116-212
C 38	13.8	69.0	1422	4	US-09-134-001C-1936
C 39	13.8	69.0	1689	4	US-09-247-155-61
C 40	13.8	69.0	4765	1	US-08-750-532-8
C 41	13.8	69.0	4765	4	US-08-894-818A-7
C 42	13.8	69.0	4765	4	US-09-445-472-5
C 43	13.8	69.0	6450	4	US-09-041-886-34
C 44	13.8	69.0	6450	4	US-08-453-998-1
C 45	13.8	69.0	38844	4	US-09-734-675-3

ALIGNMENTS

RESULT 1

US-09-053-702-3/c

; Sequence 3, Application US/09053702

; Patent No. 6229069

; GENERAL INFORMATION:

; APPLICANT: YAMADA, Shigehiro

; TITLE OF INVENTION: METHOD FOR CONTROLLING WATER CONTENT OF PLANT

; FILE REFERENCE: 230-122P

; CURRENT APPLICATION NUMBER: US/09/053,702

; CURRENT FILING DATE: 1998-04-02

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 587

; TYPE: DNA

; ORGANISM: Nicotiana tabacum

US-09-053-702-3

Query Match 77.0%; Score 15.4; DB 4; Length 587;

Best Local Similarity 94.1%; Pred. No. 29;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAC TAGTC AAAGTC 19

DB 288 AAAC TAGGACAAAGTC 272

RESULT 2

US-08-742-753-1/c

; Sequence 1, Application US/08742753

; Patent No. 5861278

; GENERAL INFORMATION:

; APPLICANT: WONG, Gordon G.

; APPLICANT: YAO, Kwok-Wing

; TITLE OF INVENTION: HNF3-delta Compositions

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA: US/08/742,753

```
/
/ FILING DATE:
/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: LAZAR, Steven R.
/ REGISTRATION NUMBER: 32,618
/ REFERENCE/DOCKET NUMBER: 5277
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 498-8260
/ TELEFAX: (617) 876-5851
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3441 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 88..2400
/
US-08-742-753-1
/
Query Match 76.0%; Score 15.2; DB 2; Length 3441;
Best Local Similarity 85.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAAAGGTCA 20
Db 2544 CATAATTAGGTCAAAGGCA 2525

RESULT 3
US-09-009-217-11
/ Sequence 11, Application US/09009217
/ Patent No. 6132729
/ GENERAL INFORMATION:
/ APPLICANT: Thorpe, Philip E.
/ APPLICANT: King, Steven W.
/ APPLICANT: Gao, Boning
/ TITLE OF INVENTION: COMBINED TISSUE FACTOR AND
/ TITLE OF INVENTION: CHEMOTHERAPUTIC METHODS AND COMPOSITIONS FOR COAGULATION
/ TITLE OF INVENTION: AND TUMOR TREATMENT
/ NUMBER OF SEQUENCES: 27
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Arnold, White & Durkee
/ STREET: P.O. Box 4433
/ CITY: Houston
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/009,217
/ FILING DATE: Concurrently Herewith
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/042,427
/ FILING DATE: 27-MAR-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/036,205
/ FILING DATE: 27-JAN-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/035,920
/ FILING DATE: 22-JAN-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hibler, David W.
/ REGISTRATION NUMBER: 41,071
/ REFERENCE/DOCKET NUMBER: UTSD:536
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 512/418-3000
/ TELEFAX: 512/418-3000
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 13865 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
US-09-009-656-11
/
Query Match 76.0%; Score 15.2; DB 3; Length 13865;
Best Local Similarity 85.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

/
/ FILING DATE: 512/474-7577
/ INFORMATION FOR SEQ ID NO: 11:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 13865 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
US-09-009-217-11
/
Query Match 76.0%; Score 15.2; DB 3; Length 13865;
Best Local Similarity 85.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAAAGGTCA 20
Db 13773 CAAATTAGGTAAGGACA 13792

RESULT 4
US-09-009-656-11
/ Sequence 11, Application US/09009656
/ Patent No. 6132730
/ GENERAL INFORMATION:
/ APPLICANT: Thorpe, Philip E.
/ APPLICANT: King, Steven W.
/ APPLICANT: Gao, Boning
/ TITLE OF INVENTION: COMBINED TISSUE FACTOR AND FACTOR VIIa
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR COAGULATION AND TUMOR
/ TITLE OF INVENTION: TREATMENT
/ NUMBER OF SEQUENCES: 27
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Arnold, White & Durkee
/ STREET: P.O. Box 4433
/ CITY: Houston
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/009,656
/ FILING DATE: Concurrently Herewith
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/042,427
/ FILING DATE: 27-MAR-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/036,205
/ FILING DATE: 27-JAN-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/035,920
/ FILING DATE: 22-JAN-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hibler, David W.
/ REGISTRATION NUMBER: 41,071
/ REFERENCE/DOCKET NUMBER: UTSD:537
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 512/418-3000
/ TELEFAX: 512/474-7577
/ INFORMATION FOR SEQ ID NO: 11:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 13865 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
US-09-009-656-11
/
Query Match 76.0%; Score 15.2; DB 3; Length 13865;
Best Local Similarity 85.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 CAAACTAGGTCAAGGTCA 20  
|||||  
Db 13773 CAAATTAGGTAAAGGACA 13792  
|||||

US-08-348-891A-1  
Query Match 76.0%; Score 15.2; DB 1; Length 15894;  
Best Local Similarity 85.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAGGTCA 20  
|||||  
Db 9920 CAAACCAAGTCAATGTCA 9901  
|||||

RESULT 6  
US-08-905-817-1/c  
; Sequence 1, Application US/08905817  
; Patent No. 5824777  
; GENERAL INFORMATION:  
; APPLICANT: SASAKI, Keiko  
; APPLICANT: MORI, Takayuki  
; APPLICANT: MAKINO, Satoshi  
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,  
; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: YOUNG & THOMPSON  
; STREET: 745 South 23rd Street  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/905,817  
; FILING DATE: 04-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/348,891  
; FILING DATE: 25-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/848,400  
; FILING DATE: 10-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 3-293625  
; FILING DATE: 14-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PATCH, Andrew J.  
; REGISTRATION NUMBER: 32,925  
; REFERENCE/DOCKET NUMBER: KP-7501A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-521-2297  
; TELEFAX: 703-685-0573  
; TELEX: 248425 EMBON  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15894 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 108..1682  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1807..3327  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 7271..9121  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3438..4442  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 9234..15782

US-08-348-891A-1/c  
; Sequence 1, Application US/08348891A  
; Patent No. 5654136  
; GENERAL INFORMATION:  
; APPLICANT: SASAKI, Keiko  
; APPLICANT: MORI, Takayuki  
; APPLICANT: MAKINO, Satoshi  
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,  
; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: YOUNG & THOMPSON  
; STREET: 745 South 23rd Street  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/348,891A  
; FILING DATE: 25-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/848,400  
; FILING DATE: 10-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 3-293625  
; FILING DATE: 14-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PATCH, Andrew J.  
; REGISTRATION NUMBER: 32,925  
; REFERENCE/DOCKET NUMBER: KP-7501  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-521-2297  
; TELEFAX: 703-685-0573  
; TELEX: 248425 EMBON  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15894 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 108..1682  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1807..3327  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3438..4442  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 5458..7107  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 7271..9121  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 9234..15782

NAME/KEY: CDS  
LOCATION: 5458..7107  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 7271..9121  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9234..15782  
US-08-905-817-1

Query Match 76.0%; Score 15.2; DB 1; Length 15894;  
Best Local Similarity 85.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAGTCA 20  
DB 9920 CAAACCCAGTCAATGTCA 9901

## RESULT 7

US-08-351-413-1  
Sequence 1, Application US/08351413  
Patent No. 5750867  
GENERAL INFORMATION:  
APPLICANT: Williams, Mark  
TITLE OF INVENTION: Maintenance of male-sterile plants  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 2046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/351,413  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/899,072  
FILING DATE: 12-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/970,849  
FILING DATE: 03-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-102PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2661 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Zea mays  
STRAIN: inbred line W-22  
PUBLICATION INFORMATION:  
AUTHORS: Hamilton et al.,  
JOURNAL: Sex Plant Reprod.

VOLUME: 2  
PAGES: 208-  
DATE: 1989  
US-08-351-413-1

Query Match 75.0%; Score 15; DB 1; Length 2661;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAA 15  
DB 1180 CAAACTAGGTCAAA 1194

## RESULT 8

US-09-025-583-1  
Sequence 1, Application US/09025583  
Patent No. 5977433  
GENERAL INFORMATION:  
APPLICANT: Williams, Mark  
TITLE OF INVENTION: Maintenance of male-sterile plants  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 2046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/025,583  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/351,413  
FILING DATE:  
APPLICATION NUMBER: US 07/899,072  
FILING DATE: 12-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/970,849  
FILING DATE: 03-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-102PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2661 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Zea mays  
STRAIN: inbred line W-22  
PUBLICATION INFORMATION:  
AUTHORS: Hamilton et al.,  
JOURNAL: Sex Plant Reprod.  
VOLUME: 2  
PAGES: 208-  
DATE: 1989

US-09-025-583-1

Query Match 75.0%; Score 15; DB 2; Length 2661;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAA 15  
|||||  
DB 1180 CAAACTAGGTCAA 1194

## RESULT 9

US-08-351-413-17/c  
; Sequence 17, Application US/08351413  
; Patent No. 5750867  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Mark  
; APPLICANT: Leemans, Jan  
; TITLE OF INVENTION: Maintenance of male-sterile plants  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; STREET: 8110 Gatehouse Road, Suite 500 East  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 2046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/351.413  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/899,072  
; FILING DATE: 12-JUN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/970,849  
; FILING DATE: 03-NOV-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svensson, Leonard R.  
; REGISTRATION NUMBER: 30,330  
; REFERENCE/DOCKET NUMBER: 2121-102PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4808 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Ecoli-HindIII fragment of plasmid pRS218  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: complement (18..401)  
; OTHER INFORMATION: /label= 3'nos  
; OTHER INFORMATION: /note= "3, regulatory sequence containing the  
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium  
; OTHER INFORMATION: T-DNA nopaline synthase gene"  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: complement (402..737)  
; OTHER INFORMATION: /label= barnase  
; OTHER INFORMATION: /note= "coding region of the barnase gene of

; OTHER INFORMATION: Bacillus amyloliquefaciens"

; FEATURE:  
; NAME/KEY: -  
; LOCATION: complement (738..1944)  
; OTHER INFORMATION: /label= PZW13  
; OTHER INFORMATION: /note= "promoter region of the Zml3 gene of Zea  
; OTHER INFORMATION: mays"  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: complement (1945..2281)  
; OTHER INFORMATION: /label= 3'nos  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: complement (2282..2554)  
; OTHER INFORMATION: /label= barstar  
; OTHER INFORMATION: /note= "coding region of the barstar gene of  
; OTHER INFORMATION: Bacillus amyloliquefaciens"  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: complement (2555..3099)  
; OTHER INFORMATION: /label= PTA29  
; OTHER INFORMATION: /note= "promoter region of the PTA29 gene of  
; OTHER INFORMATION: Nicotiana tabacum"  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 3100..3932  
; OTHER INFORMATION: /label= 35S3  
; OTHER INFORMATION: /note= "35S3" promoter sequence derived from  
; OTHER INFORMATION: cauliflower mosaic virus isolate CabbB-J1"  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 3933..4484  
; OTHER INFORMATION: /label= bar  
; OTHER INFORMATION: /note= "coding region of the phosphinothricin  
; OTHER INFORMATION: acetyltransferase gene"  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 4485..4763  
; OTHER INFORMATION: /label= 3'nos  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 2333..2356  
; OTHER INFORMATION: /label= BXOL2  
; OTHER INFORMATION: /note= "region corresponding to oligonucleotide  
; OTHER INFORMATION: BXOL2"  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: complement (2538..2586)  
; OTHER INFORMATION: /label= TA29SBXOL2  
; OTHER INFORMATION: /note= "region complementary to oligonucleotide  
; OTHER INFORMATION: TA29SBXOL2"  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: complement (2800..2823)  
; OTHER INFORMATION: /label= PTA29OL5  
; OTHER INFORMATION: /note= "region complementary to part of  
; OTHER INFORMATION: oligonucleotide PTA29OL5"  
; US-08-351-413-17  
; Query Match 75.0%; Score 15; DB 1; Length 4808;  
; Best Local Similarity 100.0%; Pred. No. 66;  
; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
; QY 1 CAAACTAGGTCAA 15  
; DB 984 CAAACTAGGTCAA 970  
; RESULT 10  
; US-09-025-583-17/c  
; Sequence 17, Application US/09025583  
; Patent No. 5977433  
; GENERAL INFORMATION:

APPLICANT: Williams, Mark  
APPLICANT: Leemans, Jan  
TITLE OF INVENTION: Maintenance of male-sterile plants  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 2046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/025,593  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/351,413  
FILING DATE:  
APPLICATION NUMBER: US 07/899,072  
FILING DATE: 12-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/970,849  
FILING DATE: 03-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-102PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4808 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: EcoRI-HindIII fragment of plasmid pTS218  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (18..401)  
OTHER INFORMATION: /label= 3'nos  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: /note= "polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA nopaline synthase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (402..737)  
OTHER INFORMATION: /label= barnase  
OTHER INFORMATION: /note= "coding region of the barnase gene of  
OTHER INFORMATION: Bacillus amyloliquefaciens"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (738..1944)  
OTHER INFORMATION: /label= PZW13  
OTHER INFORMATION: /note= "promoter region of the Zml3 gene of Zea  
OTHER INFORMATION: mays"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (1945..2281)  
OTHER INFORMATION: /label= 3'nos  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (2282..2554)

OTHER INFORMATION: /label= barstar  
OTHER INFORMATION: /note= "coding region of the barstar gene of  
OTHER INFORMATION: Bacillus amyloliquefaciens"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (2555..3099)  
OTHER INFORMATION: /label= PTA29  
OTHER INFORMATION: /note= "promoter region of the TA29 gene of  
OTHER INFORMATION: Nicotiana tabacum"  
FEATURE:  
NAME/KEY: -  
LOCATION: 3100..3932  
OTHER INFORMATION: /label= 3SS3  
OTHER INFORMATION: /note= "3SS3" promoter sequence derived from  
OTHER INFORMATION: cauliflower mosaic virus isolate CabBB-J1"  
FEATURE:  
NAME/KEY: -  
LOCATION: 3933..4484  
OTHER INFORMATION: /label= bar  
OTHER INFORMATION: /note= "coding region of the phosphinothricin  
OTHER INFORMATION: acetyltransferase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: 4485..4763  
OTHER INFORMATION: /label= 3'nos  
FEATURE:  
NAME/KEY: -  
LOCATION: 2333..2356  
OTHER INFORMATION: /label= BXOL2  
OTHER INFORMATION: /note= "region corresponding to oligonucleotide  
OTHER INFORMATION: BXOL2"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (2538..2586)  
OTHER INFORMATION: /label= TA29SBXOL2  
OTHER INFORMATION: /note= "region complementary to oligonucleotide  
OTHER INFORMATION: TA29SBXOL2"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (2800..2823)  
OTHER INFORMATION: /label= PTA29OL5  
OTHER INFORMATION: /note= "region complementary to part of  
OTHER INFORMATION: oligonucleotide PTA29OL5"  
US-09-025-583-17  
Query Match 75.0%; Score 15; DB 2; Length 4808;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CAAAACTAGGTCAAA 15  
Db 984 CAAAACTAGGTCAAA 970  
RESULT 11  
US-08-961-527-115/C  
Sequence 115, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961.527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
JS-08-961-527-115

Query Match 74.0%; Score 14.8; DB 4; Length 11303;  
Best Local Similarity 88.9%; Pred. No. 97;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2Y 1 CAATACTAGGTCAGGTC 18

Db 7784 CAATACTAGGTCAGGTC 7767

RESULT 12  
US-09-146-053-7/c  
Sequence 7, Application US/09146053A  
Patent No. 6399349  
GENERAL INFORMATION:  
APPLICANT: Ryan, James W.  
APPLICANT: Sprinkle, Terry Joe Curtis  
APPLICANT: Venema, Richard C.  
TITLE OF INVENTION: Human Aminopeptidase P Gene  
FILE REFERENCE: MCG103  
CURRENT APPLICATION NUMBER: US/09/146,053A  
CURRENT FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/057,854  
EARLIER FILING DATE: 1997-09-02  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7  
LENGTH: 16595  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-146-053-7

Query Match 74.0%; Score 14.8; DB 4; Length 16595;  
Best Local Similarity 88.9%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAATACTAGGTCAGGTC 20

Db 1472 AAATACTAGGTCAGGTC 1455

RESULT 13  
US-09-328-111-128/c  
Sequence 128, Application US/09328111  
Patent No. 6262333  
GENERAL INFORMATION:  
APPLICANT: Endese, Wilson O.  
APPLICANT: Steinmann, Kathleen E.  
APPLICANT: Astle, Jon H.  
APPLICANT: Burgess, Christopher C.  
APPLICANT: Bushnell, Steven E.

APPLICANT: Carroll III, Eddie  
APPLICANT: Catino, Theodore J.  
APPLICANT: Derti, Adnan  
APPLICANT: Ford, Donna M.  
APPLICANT: Lewis, Marcia E.  
APPLICANT: Monahan, John E.  
APPLICANT: Schlegel, Robert  
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
TITLE OF INVENTION: PRODUCTS  
FILE REFERENCE: CCD-257 (US)  
CURRENT APPLICATION NUMBER: US/09/328,111  
CURRENT FILING DATE: 1999-06-08  
EARLIER APPLICATION NUMBER: US 60/088,801  
EARLIER FILING DATE: 1998-06-10  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 128  
LENGTH: 865  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(865)  
OTHER INFORMATION: n = A,T,C or G  
US-09-328-111-128

Query Match 71.0%; Score 14.2; DB 4; Length 865;  
Best Local Similarity 80.0%; Pred. No. 1.3e+02;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATACTAGGTCAGGTC 20

Db 595 CAATACTAGGTCAGGTC 576

RESULT 14  
US-09-394-110A-3  
Sequence 3, Application US/09394110A  
Patent No. 6451594  
GENERAL INFORMATION:  
APPLICANT: Chien, Kenneth  
APPLICANT: Wang, Yibin  
APPLICANT: Evans, Sylvia  
TITLE OF INVENTION: No. 6451594el Recombinant Adenovirus for Tissue Specific Express  
FILE REFERENCE: 6627-PA8045  
CURRENT APPLICATION NUMBER: US/09/394,110A  
CURRENT FILING DATE: 1999-09-10  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3  
LENGTH: 1026  
TYPE: DNA  
ORGANISM: Mus musculus  
US-09-394-110A-3

Query Match 71.0%; Score 14.2; DB 4; Length 1026;  
Best Local Similarity 84.2%; Pred. No. 1.3e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATACTAGGTCAGGTC 20

Db 233 AAATACTAGGTCAGGTC 251

RESULT 15  
US-09-724-864-16  
Sequence 16, Application US/09724864  
Patent No. 6380362  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Murison, James G.  
TITLE OF INVENTION: Polynucleotides, polypeptides expressed  
by the polynucleotides and methods for their use.



Search completed: February 20, 2003, 01:12:47  
Job time : 17.5559 secs

Result No.	Score	Query #		Length	DB	ID	Description
		Match					
C 1	18.4	92.0	301	9	AV165535	AV165535	AV165535
C 2	18.4	92.0	307	10	AW457734	AW457734	UI-M-BH3-
C 3	18.4	92.0	429	10	EB690786	EB690786	UI-M-BH3-
C 4	18.4	92.0	453	12	BF470083	BF470083	UI-M-BH3-
C 5	18.4	92.0	454	10	AW493749	AW493749	UI-M-BH3-
C 6	18.4	92.0	454	10	AW494112	AW494112	UI-M-BH3-

Email: genome-res@rtc.riken.go.jp  
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

# FEATURES

## source

Location/Qualifiers  
1. .307  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH3-ari-f-01-0-UI"  
/clone\_lib="NIH\_BMAP\_M\_S4"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/notes="Vector: pT7T3D-Pac (Pharmacia)  
polylinker; Site 1: Not I; Site 2: Eco RI; The  
NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
ultimately derived from a mixture of individually tagged  
normalized libraries from ten regions of the mouse brain  
(cerebellum, brain stems, olfactory bulbs, hypothalamus,  
cortex, amygdala, basal ganglia, pineal gland, striatum,  
hippocampus) after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: NIH\_BMAP\_M\_S4,  
NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
was used as a driver in a hybridization with a pool of  
the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
libraries in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library)  
was purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
NIH\_BMAP\_M\_S4 library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)  
TAG\_LIB=NIH\_BMAP\_M\_S4  
TAG\_TISSUE=Olfactory-bulbs  
TAG\_SEQ=CATGG"

BASE COUNT 96 a 52 c 53 g 100 t

ORIGIN  
Query Match 92.0%; Score 18.4; DB 9; Length 301;  
Best Local Similarity 95.0%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CAAAAGTGGTCAAAAGGTCA 20  
Db 21 CAAAAGTGGTCAAAAGGTCA 2

RESULT 2  
AW457734 307 bp mRNA linear EST 24-FEB-2000  
LOCUS  
DEFINITION  
UI-M-BH3-ari-f-01-0-UI.s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
UI-M-BH3-ari-f-01-0-UI 3', mRNA sequence.  
ACCESSION  
AW457734  
VERSION  
AW457734.1 GI:7027951  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 307)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: m88@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
normalized olfactory bulbs library cDNA Library Preparation: M.B.  
Soares Lab Clome distribution: Researchers may obtain BMAP cDNA

clones from RESEARCH GENETICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=Yes.

Location/Qualifiers  
1. .307  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH3-ari-f-01-0-UI"  
/clone\_lib="NIH\_BMAP\_M\_S4"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/notes="Vector: pT7T3D-Pac (Pharmacia)  
polylinker; Site 1: Not I; Site 2: Eco RI; The  
NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
ultimately derived from a mixture of individually tagged  
normalized libraries from ten regions of the mouse brain  
(cerebellum, brain stems, olfactory bulbs, hypothalamus,  
cortex, amygdala, basal ganglia, pineal gland, striatum,  
hippocampus) after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: NIH\_BMAP\_M\_S4,  
NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
was used as a driver in a hybridization with a pool of  
the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
libraries in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library)  
was purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
NIH\_BMAP\_M\_S4 library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)  
TAG\_LIB=NIH\_BMAP\_M\_S4  
TAG\_TISSUE=Olfactory-bulbs  
TAG\_SEQ=CATGG"

BASE COUNT 96 a 52 c 53 g 100 t

ORIGIN  
Query Match 92.0%; Score 18.4; DB 10; Length 307;  
Best Local Similarity 95.0%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAAGTGGTCAAAAGGTCA 20  
Db 286 CAAAAGTGGTCAAAAGGTCA 305

RESULT 3  
BB690786/c 429 bp mRNA linear EST 10-OCT-2001  
LOCUS  
DEFINITION  
BB690786 RIKEN full-length enriched, 12 days embryo female  
mullerian duct Mus musculus cDNA clone 6820449M04 3', mRNA  
sequence.  
BB690786  
BB690786  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 429)  
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imcanta, K., Ishii  
Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,

clones from RESEARCH GENETICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=Yes.

Location/Qualifiers  
1. .307  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH3-ari-f-01-0-UI"  
/clone\_lib="NIH\_BMAP\_M\_S4"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/notes="Vector: pT7T3D-Pac (Pharmacia)  
polylinker; Site 1: Not I; Site 2: Eco RI; The  
NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
ultimately derived from a mixture of individually tagged  
normalized libraries from ten regions of the mouse brain  
(cerebellum, brain stems, olfactory bulbs, hypothalamus,  
cortex, amygdala, basal ganglia, pineal gland, striatum,  
hippocampus) after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: NIH\_BMAP\_M\_S4,  
NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
was used as a driver in a hybridization with a pool of  
the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
libraries in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library)  
was purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
NIH\_BMAP\_M\_S4 library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)  
TAG\_LIB=NIH\_BMAP\_M\_S4  
TAG\_TISSUE=Olfactory-bulbs  
TAG\_SEQ=CATGG"

BASE COUNT 96 a 52 c 53 g 100 t

ORIGIN  
Query Match 92.0%; Score 18.4; DB 10; Length 307;  
Best Local Similarity 95.0%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAAGTGGTCAAAAGGTCA 20  
Db 286 CAAAAGTGGTCAAAAGGTCA 305

RESULT 3  
BB690786/c 429 bp mRNA linear EST 10-OCT-2001  
LOCUS  
DEFINITION  
BB690786 RIKEN full-length enriched, 12 days embryo female  
mullerian duct Mus musculus cDNA clone 6820449M04 3', mRNA  
sequence.  
BB690786  
BB690786  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 429)  
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imcanta, K., Ishii  
Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,



Query Match 92.0%; Score 18.4; DB 12; Length 453;  
 Best Local Similarity 95.0%; Pred. No. 2.6e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAACTAGTCAAGGTCA 20  
 DB 360 CAAACTAGTCAAGGTCA 341

RESULT 5  
 AW493749  
 LOCUS AW493749 454 bp mRNA linear EST 24-FEB-2000  
 DEFINITION UI-M-BH3-auc-g-04-0-UI.s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
 ACCESSION AW493749  
 VERSION AW493749.1 GI:7064030  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 454)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704477  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mEST@mail.nih.gov  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 normalized olfactory bulbs library cDNA Library Preparation: M.B.  
 Soares Lab Clones distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENETICS. It should be noted that Bento Soares  
 is generating a small number of additional specialized  
 non-redundant arrays of BMAP cDNAs whose availability will be  
 considered under appropriate and limited collaborative arrangements  
 Seg primer: M13 Forward  
 POLYA=Yes

FEATURES  
 source  
 1..454  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-auc-g-04-0-UI"  
 /clone\_lib="NIH\_BMAP\_M\_S4"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note=vector: pRT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; The  
 NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
 ultimately derived from a mixture of individually tagged,  
 normalized libraries from ten regions of the mouse brain  
 (cerebellum, brain stems, olfactory bulbs, hypothalamus,  
 cortex, amygdala, basal ganglia, pineal gland, striatum,  
 hippocampus) after a series of subtractions to reduce the  
 representation of cDNAs from which ESTs had already been  
 generated. The following serially subtracted libraries  
 were generated in this process: NIH\_BMAP\_M\_S4,  
 NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
 NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
 (NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
 cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
 NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived

was used as a driver in a hybridization with a pool of  
 the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
 libraries in the form of single-stranded circles. The  
 remaining single-stranded circles (subtracted library)  
 was purified by hydroxyapatite column chromatography,  
 converted to double-stranded circles and electroporated  
 into DH10B bacteria (lifetechnologies) to generate the  
 NIH\_BMAP\_M\_S4 library. This procedure has been previously  
 described (Bonaldo, Lennon and Soares, Genome Research  
 6:791-806, 1996)  
 TAG LIB=NIH\_BMAP\_M\_S4  
 TAG\_TISSUE=Olfactory-bulbs  
 TAG\_SEQ-CATGG" 96 g 130 t  
 BASE COUNT 153 a 75 c 96 g 130 t  
 ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 454;  
 Best Local Similarity 95.0%; Pred. No. 2.6e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAACTAGTCAAGGTCA 20  
 DB 286 CAAACTAGTCAAGGTCA 305

RESULT 6  
 AW494112  
 LOCUS AW494112 454 bp mRNA linear EST 24-FEB-2000  
 DEFINITION UI-M-BH3-aui-g-09-0-UI.s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
 ACCESSION AW494112  
 VERSION AW494112.1 GI:7064393  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 454)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704477  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mEST@mail.nih.gov  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 normalized olfactory bulbs library cDNA Library Preparation: M.B.  
 Soares Lab Clones distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENETICS. It should be noted that Bento Soares  
 is generating a small number of additional specialized  
 non-redundant arrays of BMAP cDNAs whose availability will be  
 considered under appropriate and limited collaborative arrangements  
 Seg primer: M13 Forward  
 POLYA=Yes

FEATURES  
 source  
 1..454  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-aui-g-09-0-UI"  
 /clone\_lib="NIH\_BMAP\_M\_S4"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note=vector: pRT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not 1; Site 2: Eco RI; The NIH BMAP\_M\_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH BMAP\_M\_S4, NIH BMAP\_M\_S3.3, NIH BMAP\_M\_S3.2, NIH BMAP\_M\_S3.1, NIH BMAP\_M\_S2, NIH BMAP\_M\_S1. The subtracted library (NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified cDNA inserts from NIH BMAP\_M\_S3.3, NIH BMAP\_M\_S3.2, and NIH BMAP\_M\_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH BMAP\_M\_S3.3, NIH BMAP\_M\_S3.2, and NIH BMAP\_M\_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH BMAP\_M\_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG\_LIB=NIH\_BMAP\_M\_S4  
TAG\_TISSUE=olfactory-bulbs  
TAG\_SEQ=CATGG

BASE COUNT 152 a 76 c 97 g 129 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 454;  
Best Local Similarity 95.0%; Pred. No. 2.6e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCACAAAGGTCA 20  
|||||  
Db 286 CAAACTAGATCAAAAGGTCA 305

RESULT 7  
BM894724/c  
LOCUS BM894724 469 bp mRNA linear EST 28-MAR-2002  
DEFINITION cDNA clone IMAGE:5681858 5', mRNA sequence.  
ACCESSION BM894724  
VERSION BM894724.1 GI:19350192  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 469)

REFERENCE  
AUTHORS Melton, D., Brown, J., Keny, G., Permutt, A., Lee, C., Kaestner, K., Leshchinskiy, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Eilstein, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T., Jackson, Y., and Bowers, Y.  
TITLE Endocrine Pancreas Consortium  
JOURNAL Unpublished (2000)  
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@bioph.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center This clone is available royalty-free through LLNL; please contact the IMAGE consortium (info@image.llnl.gov) for further information

Seq primer: -4ORP from Gibco  
High quality sequence stop: 432.  
Location/Qualifiers  
1. .469  
/organism="Mus musculus"  
/strain="ICR"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5681858"  
/clone\_lib="Melton Mouse E16 5 Pancreas Library 2 M16B2"  
/sex="Both"  
/tissue\_type="Total pancreas"  
/dev\_stage="Embryonic day 16.5"  
/lab\_host="TOP10"  
/notes="Organ: Pancreas; Vector: pBluescript II SK; Site 1: Not1; Site 2: SalI; Library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming size-selected by column fractionation; average insert size 1.06kb. Primary library, unamplified."

BASE COUNT 115 a 110 c 81 g 163 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 469;  
Best Local Similarity 95.0%; Pred. No. 2.6e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCACAAAGGTCA 20  
|||||  
Db 252 CAAACTAGATCAAAAGGTCA 233

RESULT 8  
AA144171/c  
LOCUS AA144171 477 bp mRNA linear EST 18-FEB-1997  
DEFINITION mq54g11.r1 Soares thymus\_2NBMT Mus musculus cDNA clone IMAGE:582596  
5', mRNA sequence.  
ACCESSION AA144171  
VERSION AA144171.1 GI:1713539  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 477)

REFERENCE  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:357244

Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 420.  
Location/Qualifiers  
1. .477

FEATURES  
source  
/organism="Mus musculus"  
/strain="CS7BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:582596"  
/clone\_lib="Soares thymus\_2NBMT"  
/sex="male"  
/tissue\_type="Thymus"  
/dev\_stage="4 weeks"

```

/lab host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5,
TGTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT      129 a   108 c   84 g   156 t
ORIGIN

```

```

Query Match      92.0%; Score 18.4; DB 9; Length 477;
Best Local Similarity 95.0%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 CAAAAGTGGTCAAAAGTCA 20
|||||
Db 238 CAAAAGTGGTCAAAAGTCA 219
|||||

```

```

RESULT 9
BB754909/c
LOCUS BB754909 RIKEN full-length enriched, melanocyte Mus musculus cDNA
DEFINITION clone G270045K23 3', mRNA sequence.
ACCESSION BB754909
VERSION BB754909.1 GI:16184944
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 500)
Akimura T., Arakawa T., Carninci P., Furuno M., Hanagaki T.,
Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T., Imofani K., Iehii
I., Ito M., Kawai J., Kojima Y., Konno H., Kouda M., Matsuyama T.,
Nakamura M., Nishi K., Nomura K., Numasaki R., Okazaki Y., Okido T.,
Saito R., Sakai C., Sakai K., Sakazume N., Sasaki D., Sato K.,
Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Suzuki H., Tagawa
A., Takahashi F., Takaku-Akahira S., Tanaka T., Tomaru A., Toya T.,
Watabiki A., Yasunishi A., Muramatsu M. and Hayashizaki Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura T., et al.
2001)

```

```

JOURNAL
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh
M., Konno H., Okazaki Y., Muramatsu M. and Hayashizaki Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E.,
Watabiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura
S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A. and
Hayashizaki Y.

```

```

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1751-1771 (2000)
Konno H., Fukunishi Y., Shibata K., Itoh M., Carninci P., Sugahara
Y. and Hayashizaki Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

```

```

Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

```

```

FEATURES
source
1..500
/organism="Mus musculus"
/strain="C57BL/6J"
/db xref="taxon:10090"
/clone="G270045K23"
/clone_lib="RIKEN full-length enriched, melanocyte"
/cell_type="melanocyte"
/notes="pooled tissues ; (tissue_type=cerebellum,
dev_stage=16 days neonate, sex=mixed),
(tissue_type=cerebellum, dev_stage=0 day neonate,
sex=mixed), (tissue_type=hippocampus, dev_stage=9 days
sex=male), (tissue_type=whole body, dev_stage=9 days
embryo, sex=mixed); (tissue_type=lung, dev_stage=13 days
embryo, sex=mixed);"

```

```

BASE COUNT      134 a   115 c   84 g   167 t
ORIGIN

```

```

Query Match      92.0%; Score 18.4; DB 10; Length 500;
Best Local Similarity 95.0%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 CAAAAGTGGTCAAAAGTCA 20
|||||
Db 218 CAAAAGTGGTCAAAAGTCA 199
|||||

```

```

RESULT 10
BE449139/c
LOCUS BE449139 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:3331555
DEFINITION 5', mRNA sequence.
ACCESSION BE449139
VERSION BE449139.1 GI:9448716
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 503)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

```

```

AUTHORS
MGI:1075719
Seq primer: -40RP from Gibco
High quality sequence stop: 467.

```

```

FEATURES
source
1..503
/organism="Mus musculus"
/strain="C57BL/6J"
/db xref="taxon:10090"
/clone="IMAGE:3331555"
/clone_lib="Soares mouse 3NDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5,
TGTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through

```

three rounds of normalization, and was constructed by

BASE COUNT 160 a 82 c 86 g 175 t  
 ORIGIN  
 Query Match 92.0%; Score 18.4; DB 10; Length 503;  
 Best Local Similarity 95.0%; Pred. No. 2.7e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCARAGGTCA 20  
 Db 48 CAAACTAGGTCARAGGTCA 29

RESULT 11  
 LOCUS AA276952/c 517 bp mRNA linear EST 01-APR-1997  
 DEFINITION vc42510.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:777282  
 5', mRNA sequence.  
 ACCESSION AA276952  
 VERSION AA276952.1 GI:1919597  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 517)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lemon, G., Soares, B., Wilson, R. and  
 Watson, R.

TITLE The WashU-RHMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-RHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMGI Consortium (info@imgi.llnl.gov) for further information.  
 MGI:470138

Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 475.

FEATURES  
 source  
 Location/Qualifiers  
 1..517  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:777282"  
 /clone\_lib="Soares mouse 3NbMS"  
 /sex="male"  
 /tissue\_type="Spleen"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Vector: pTT73-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5'  
 TGTTCACATGAGTGGAGCGCGCGCTGTGTGTGTGTGTGTGTGTGT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pTT73 vector. RNA  
 provided by Dr. Bertrand Jordan. Library went through  
 three rounds of normalization, and was constructed by  
 Bento Soares and M.Fatima Bonaldo."  
 144 a 106 c 88 g 179 t

BASE COUNT 144 a 106 c 88 g 179 t  
 ORIGIN  
 Query Match 92.0%; Score 18.4; DB 9; Length 517;  
 Best Local Similarity 95.0%; Pred. No. 2.8e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCARAGGTCA 20  
 Db 171 CAAACTAGGTCARAGGTCA 152

RESULT 12  
 LOCUS AW492996  
 DEFINITION UI-M-BH3-aty-c-01-0-UI.s1 NIH BMAP\_M\_S4 Mus musculus cDNA clone  
 UI-M-BH3-aty-c-01-0-UI 3', mRNA sequence.  
 ACCESSION AW492996  
 VERSION AW492996.1 GI:7063277  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 557)  
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704477  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9830  
 Email: mst@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 normalized olfactory bulbs library cDNA library Preparation: M.B.  
 Soares Lab Clones distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENETICS. It should be noted that Bento Soares  
 is generating a small number of additional specialized  
 non-redundant arrays of BMAP cDNAs whose availability will be  
 considered under appropriate and limited collaborative arrangements  
 Seq primer: M13 Forward  
 POLYA=Yes.

FEATURES  
 source  
 Location/Qualifiers  
 1..557  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-aty-c-01-0-UI"  
 /clone\_lib="NIH BMAP\_M\_S4"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="vector: pTT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; The  
 NIH BMAP\_M\_S4 library is a subtracted library of a series,  
 ultimately derived from a mixture of individually tagged  
 normalized libraries from ten regions of the mouse brain  
 (cerebellum, brain stems, olfactory bulbs, hypothalamus,  
 cortex, amygdala, basal ganglia, pineal gland, striatum,  
 hippocampus) after a series of subtractions to reduce the  
 representation of cDNAs from which ESTs had already been  
 generated. The following serially subtracted libraries  
 were generated in this process: NIH BMAP\_M\_S4,  
 NIH BMAP\_M\_S3.3, NIH BMAP\_M\_S3.2, NIH BMAP\_M\_S3.1,  
 NIH BMAP\_M\_S2, NIH BMAP\_M\_S1. The subtracted library  
 (NIH BMAP\_M\_S4) was constructed as follows: PCR amplified  
 cDNA inserts from NIH BMAP\_M\_S3.3, NIH BMAP\_M\_S3.2, and  
 NIH BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
 was used as a driver in a hybridization with a pool of  
 the NIH BMAP\_M\_S3.3, NIH BMAP\_M\_S3.2, and NIH BMAP\_M\_S3.1  
 libraries in the form of single-stranded circles. The  
 remaining single-stranded circles (subtracted library)



was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH BMAP M.S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996).

TAG LIB=NIH\_BMAP\_M.S4  
TAG\_TISSUE=olfactory-bulbs  
TAG\_SEQ=CATGG

BASE COUNT 184 a 123 g 153 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 557;  
Best Local Similarity 95.0%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAACCTAGGTCAAAGGTCA 20  
|||||  
Db 286 CAAAACCTAGGTCAAAGGTCA 305

RESULT 13  
LOCUS AAB59634  
DEFINITION UI-R-E0-bs-h-10-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone  
binding protein Etr-3 mRNA, complete cds, mRNA sequence.  
ACCESSION AAB59634  
VERSION AAB59634.1 GI:4230179  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 573)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 9704477  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949154.  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.wiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 18-Day-Embryo library. cDNA library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1777095  
Seq primer: M13 Forward  
POLYA=No.  
FEATURES  
source  
Location/Qualifiers  
1..573  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-E0-bs-h-10-0-UI"  
/dev\_stage="embryonic"  
/note="Vector: p773B-Pac (Life Technologies)"  
/lab host="DH10B (Life Technologies)"  
polylinker: Site 1: NotI; Site 2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows

identification of the library of origin of a clone within the mixture."

BASE COUNT 121 a 148 c 98 g 206 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 573;  
Best Local Similarity 95.0%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAACCTAGGTCAAAGGTCA 20  
|||||  
Db 494 CAAAACCTAGGTCAAAGGTCA 475

RESULT 14  
LOCUS BG347085/c

DEFINITION BG347085 575 bp mRNA linear EST 28-FEB-2001  
IMAGE:4437258 5', mRNA sequence.

ACCESSION BG347085  
VERSION BG347085.1 GI:13167509  
KEYWORDS EST.  
SOURCE African clawed frog.

ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 575)  
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hallier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.  
WashU Xenopus EST project, 1999  
UNPUBLISHED (1999)  
CONTACT: Sandy Clifton, Ph.D.  
WashU Xenopus EST project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov  
Seq primer: -40RP from Gibco  
High quality sequence stop: 472.

FEATURES  
source  
Location/Qualifiers  
1..575  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="IMAGE:4437258"  
/clone lib="Wellcome CRC pcDNA1 St24-26"  
/tissue type="pooled embryos, stage 24-26"  
/lab host="DH10B (phage-resistant)"  
/note="Vector: pcDNA1; Site 1: NotI; Site 2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Newkooop and Faber. Library was constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute)."  
BASE COUNT 151 a 126 c 112 g 186 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 575;  
Best Local Similarity 95.0%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAACCTAGGTCAAAGGTCA 20  
|||||  
Db 326 CAAAACCTAGGTCAAAGGTCA 307

RESULT 15

BM238433 606 bp mRNA linear EST 31-JAN-2002  
K0519A09-3 NTA Mouse Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1+)  
CDNA Library (Long) Mus musculus cDNA clone K0519A09 3', mRNA  
sequence.  
ACCESSION BM238433  
VERSION  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 606)  
Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G.,  
Taub, D., Longo, D.L., Keller, J. and Ko, M.S.H.  
Systematic Analyses of NIA Mouse Hematopoietic Stem Cell  
(Lin-/c-Kit+/Sca-1+) cDNA Library (Long)  
Unpublished (2001)  
Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
Plate: K0519 row: A column: 09  
Seq primer: -21M13 Forward  
High quality sequence stop: 606  
POLYA=Yes.

FEATURES

Location/Qualifiers  
1..606  
/organism="Mus musculus"  
/strain="C57BL/6Ncr"  
/db\_xref="niaEST:K0519A09-3"  
/db\_xref="taxon:10090"  
/clone="K0519A09"  
/clone\_lib="NIA Mouse Hematopoietic Stem Cell  
(Lin-/c-Kit+/Sca-1+) cDNA Library (Long)"  
/issue\_type="Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1+)"  
/dev\_stages="Age approx. 10 weeks old"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:  
NotI; Mouse cDNA project by the Laboratory of Genetics,  
National Institute on Aging (NIA), Intramural Research  
Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is  
a long-transcript enriched cDNA library (Ref. Genome Res.  
11: 1553-1558 (2001). [PMID: 1154199]). Total RNAs were  
obtained from Drs. Dennis Taub, Pan Longo (National  
Institute on Aging, USA), Jonathan Keller (National Cancer  
Institute, USA). Double-stranded cDNAs were synthesized  
with an Oligo(dT) primer (Invitrogen).  
5'-pGACTAGTCTAGATCGGAGCGCGCCCTTTT-TTTT-3' from  
4.8 ug of total RNA, treated with T4 DNA polymerase, and  
purified by ethanol-precipitation. The cDNAs were ligated  
to Lona-linker LL-Sal4, purified by phenol/chloroform, and  
separated from free linkers by Centricon 100. Then, the  
cDNAs were amplified by long-range high fidelity PCR using  
Ex Taq polymerase (Takara) with a primer Sal4-S. The  
products were purified by phenol/chloroform and Centricon  
100. The cDNAs were digested with SalI and NotI enzymes  
and cloned into SalI/NotI site of pSPORT1 plasmid vector.  
The DH10B E. coli host was transformed with the ligation  
mixture by the standard chemical method. The average  
insert size is about 2.7 kb. The library was constructed  
by Yulan Piao (NIA)."

BASE COUNT 195 a 113 c 134 g 164 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 13; Length 606;  
Best Local Similarity 95.0%; Pred. No. 3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAACTAGTCAAGGTCA 20

Db 283 CAAACTAGTCAAGGTCA 302

Search completed: February 20, 2003, 01:08:46  
Job time : 101.314 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:28:33 ; Search time 118.185 seconds  
(without alignments)  
9357.426 Million cell updates/sec

Title: US-09-808-388-2

Perfect score: 38

Sequence: 1 caaaactaggtcaaaaggtcaaaactagggtcaaaagggtca 38

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.hcg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pr.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.hcg.hum.\*

31: em.hcg.inv.\*

32: em.hcg.other.\*

33: em.hcg.mus.\*

34: em.hcg.pln.\*

35: em.hcg.rod.\*

36: em.hcg.nam.\*

37: em.hcg.vrt.\*

38: em.sy.\*

39: em.hugo.hum.\*

40: em.hugo.mus.\*

41: em.hugo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	38	100.0	38	6	AX251574	AX251574 Sequence
2	26.4	69.5	1634	10	AF169013	AF169013 Rattus no
3	26.4	69.5	3587	10	BC026856	BC026856 Mus muscu
4	26.4	69.5	5446	10	AF090597	AF090597 Mus muscu
5	26.4	69.5	5546	10	AF090696	AF090696 Mus muscu
6	26.4	69.5	112269	2	AL845515	AL845515 Mus muscu
7	26.4	69.5	225433	2	AC125485	AC125485 Mus muscu
8	26.4	69.5	245916	2	AL845492	AL845492 Mus muscu
9	26	68.4	210768	2	AC107711	AC107711 Mus muscu
10	25	65.8	41	6	AX251575	AX251575 Sequence
11	25	65.8	332	6	AX251578	AX251578 Sequence
12	24	63.2	101913	2	AC131487	AC131487 Lytechinu
13	24	62.6	153172	2	AC126043	AC126043 Mus muscu
14	23.8	62.1	2113	9	HSU63289	U63289 Human RNA-b
15	23.6	62.1	2205	9	BC031079	BC031079 Homc sapi
16	23.6	62.1	2926	9	AK054655	AK054655 Homc sapi
17	23.6	62.1	2926	9	AK054655	AK054655 Homc sapi
18	23.6	62.1	5941	14	AF036333	AF036333 Diatraea
19	23.6	62.1	5941	14	AF036333	AF036333 Diatraea
20	23.6	62.1	71468	2	AC102077	AC102077 Mus muscu
21	23.6	62.1	101913	2	AC131487	AC131487 Lytechinu
22	23.6	62.1	169406	2	AC090582	AC090582 Homo sapi
23	23.6	62.1	183556	2	AC019059	AC019059 Homo sapi
24	23.6	62.1	197925	10	AL672241	AL672241 Mouse DNA
25	23.6	62.1	202339	9	AC090559	AC090559 Homo sapi
26	23.6	62.1	250352	2	AC016982	AC016982 Mus muscu
27	23.6	62.1	251891	2	AC084742	AC084742 Mus muscu
28	23.4	61.6	165449	2	AC102319	AC102319 Mus muscu
29	23.4	61.6	190155	2	AC128411	AC128411 Rattus no
30	23.2	61.1	64	6	AX063387	AX063387 Sequence
31	23.2	61.1	154494	2	AP005070	AP005070 Oryza sat
32	23.2	61.1	155818	9	AC073504	AC073504 Homo sapi
33	23.2	61.1	193389	2	AL645531	AL645531 Mus muscu
34	23.2	61.1	203613	2	AC118575	AC118575 Lemtur cat
35	23.2	61.1	210967	2	AC121814	AC121814 Mus muscu
36	23.2	61.1	221194	2	AL627122	AL627122 Mus muscu
37	22.8	60.0	64827	8	AB016873	AB016873 Arabidops
38	22.8	60.0	79804	5	AC113579	AC113579 Tetraodon
39	22.8	60.0	115385	5	AC113581	AC113581 Tetraodon
40	22.8	60.0	191976	2	AC110411	AC110411 Rattus no
41	22.6	59.5	731	6	E03296	E03296 cDNA sequen
42	22.6	59.5	6455	5	XLAI31094	XLAI31094 Xenopus l
43	22.6	59.5	154998	2	AC109787	AC109787 Bos tauru
44	22.6	59.5	158374	2	AC120953	AC120953 Rattus no
45	22.6	59.5	202375	2	AC109798	AC109798 Bos tauru

ALIGNMENTS

RESULT 1  
AX251574  
LOCUS AX251574 38 bp DNA linear PAT 05-OCT-2001  
DEFINITION Sequence 2 from Patent WO0168845.  
ACCESSION AX251574  
VERSION AX251574.1 GI:15984997  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequences.  
REFERENCE 1 (Bases 1 to 38)  
AUTHORS Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bersziat,G.  
TITLE Inflammation-inducible hybrid promoters, vectors containing same  
and uses thereof  
JOURNAL Patent: WO 0168845-A 2 20-SEP-2001;



## KEYWORDS

## SOURCE ORGANISM

Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 5446)  
Choi, D.K., Ito, T., Tsukahara, F., Hirai, M. and Sakaki, Y.  
Developmentally-regulated expression of mNapor encoding an apoptosis-induced ELAV-type RNA binding protein

JOURNAL MEDLINE  
PUBMED  
Gene 237 (1), 135-142 (1999)

## REFERENCE

2 (bases 1 to 5446)  
Choi, D.K., Ito, T. and Sakaki, Y.  
Characterization and spatial distribution of the mNapor during murine embryogenesis

## JOURNAL

3 (bases 1 to 5446)  
Choi, D.K., Ito, T. and Sakaki, Y.  
Direct Submission

JOURNAL  
Submitted (08-SEP-1998) Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan

## FEATURES

## source

Location/Qualifiers  
1..5446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="A2-A3"  
/tissue\_type="brain"  
/dev\_stage="fetus"  
1..5446  
/gene="Napor-3"  
66..1592  
/gene="Napor-3"  
/note="contains three copies of evolutionarily conserved RNA recognition motif; developmentally regulated"

## gene

## CDS

/product="apoptosis-related RNA binding protein"  
/codon\_start=1  
/protein\_id="A013764.1"  
/db\_xref="GI:4249674"  
/translation="MRCPSAVTWNRELLISNGTANKMGALDHSDDPPDAIKMFPVGIIPSWSEKELKEFPIGAVYQINVLDRSQNPPOSKGCCFTVYTRKAALEAQNALNKLTPGMHPIQMKPADSEKSLNKLTPGMHPIQMKPADSEKNAVEDKLFIMVSKKCNENDIRVMFSPFGQIEEHLRILPGDLSRGCAFTVSTRAMONAKAMHOSQTMGCGSPPIVVKFADTQKKEQRLQQLAQOQLTATGNLTGLGLTPYLLAQATSSNLGAFSGIQAGVGNALQQLATLAATAAATATSTNANPLSTSSALGALTSPVAASTPNTAGAAWNSLTSGLTGLAGATVGLNNINAGMALNGSLGALTGTNGTACTMDALTQAYSGIQYAAALPTLYSQSLQQSAGSKCEGEGANLFIHLPGFQDILQMFEPGNIYAAKVFIDKQTNLSKCFGVSDNPVSAQAALQAMNGFQIGMKRLKVLKRSKNDSPY"

BASE COUNT 1511 a 1133 c 1094 g 1708 t

## ORIGIN

Query Match 69.5%; Score 26.4; DB 10; Length 5446;  
Best Local Similarity 96.4%; Pred. No. 19;  
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 TCAAGGTCAAAAGTCTAGGTCAAAGGTCA 38

Db 1788 TCAAGGTCAAAAGTCTAGGTCAAAGGTCA 1761

## RESULT 5

## AF090696/c

## LOCUS

DEFINITION Mus musculus apoptosis-related RNA binding protein (Napor-1) mRNA, complete cds.

## ACCESSION

## AF090696

## VERSION

## SOURCE

## ORGANISM

AF090696 5546 bp mRNA linear ROD 22-OCT-1999

GI:4249671

1 (bases 1 to 112269)

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 5546)  
Choi, D.K., Ito, T., Tsukahara, F., Hirai, M. and Sakaki, Y.  
Developmentally-regulated expression of mNapor encoding an apoptosis-induced ELAV-type RNA binding protein

JOURNAL MEDLINE  
PUBMED  
Gene 237 (1), 135-142 (1999)

## REFERENCE

2 (bases 1 to 5546)  
Choi, D.K., Ito, T. and Sakaki, Y.  
Characterization and spatial distribution of the mNapor during murine embryogenesis

## JOURNAL

3 (bases 1 to 5546)  
Choi, D.K., Ito, T. and Sakaki, Y.  
Direct Submission

JOURNAL  
Submitted (08-SEP-1998) Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan

## FEATURES

## source

Location/Qualifiers  
1..5546  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="A2-A3"  
/tissue\_type="brain"  
/dev\_stage="fetus"  
1..5546  
/gene="Napor-1"  
238..1692  
/gene="Napor-1"  
/note="contains three copies of evolutionarily conserved RNA recognition motif; developmentally regulated"

## gene

## CDS

/product="apoptosis-related RNA binding protein"  
/codon\_start=1  
/protein\_id="A013763.1"  
/db\_xref="GI:4249672"  
/translation="MNGALDHSDDPPDAIKMFPVGIIPSWSEKELKEFPIGAVYQINVLDRSQNPPOSKGCCFTVYTRKAALEAQNALNKLTPGMHPIQMKPADSEKNAVEDKLFIMVSKKCNENDIRVMFSPFGQIEEHLRILPGDLSRGCAFTVSTRAMONAKAMHOSQTMGCGSPPIVVKFADTQKKEQRLQQLAQOQLTATGNLTGLGLTPYLLAQATSSNLGAFSGIQAGVGNALQQLATLAATAAATATSTNANPLSTSSALGALTSPVAASTPNTAGAAWNSLTSGLTGLAGATVGLNNINAGMALNGSLGALTGTNGTACTMDALTQAYSGIQYAAALPTLYSQSLQQSAGSKCEGEGANLFIHLPGFQDILQMFEPGNIYAAKVFIDKQTNLSKCFGVSDNPVSAQAALQAMNGFQIGMKRLKVLKRSKNDSPY"

BASE COUNT 1540 a 1176 c 1108 g 1722 t

## ORIGIN

Query Match 69.5%; Score 26.4; DB 10; Length 5546;  
Best Local Similarity 96.4%; Pred. No. 19;  
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 TCAAGGTCAAAAGTCTAGGTCAAAGGTCA 38

Db 1888 TCAAGGTCAAAAGTCTAGGTCAAAGGTCA 1861

## RESULT 6

## AL845515

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

AL845515 112269 bp DNA linear HTG 16-AUG-2002  
Mus musculus chromosome 2 clone RP23-256D19, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 31 unordered pieces.

AL845515.1 GI:22316245

HTG; HTGS PHASE1.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 112269)

Burton, J.

TITLE  
JOURNAL

Direct Submission  
Submitted (14-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

## COMMENT

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: BM256D19  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 101832 bases at least Q40  
Consensus quality: 105851 bases at least Q30  
Consensus quality: 107942 bases at least Q20  
Insert size: 109269; sum-of-contigs  
Insert size: 186792; 1.2% error; agarose-fp  
Quality coverage: 2.89x in Q20 bases; sum-of-contigs Quality  
coverage: 2.22x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 31 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2026: contig of 2026 bp in length  
\* 2027 2126: gap of 100 bp  
\* 2127 4515: contig of 2389 bp in length  
\* 4516 4615: gap of 100 bp  
\* 4616 7195: contig of 2580 bp in length  
\* 7196 7295: gap of 100 bp  
\* 7296 14681: contig of 7386 bp in length  
\* 14682 14781: gap of 100 bp  
\* 14782 21050: contig of 6269 bp in length  
\* 21051 21150: gap of 100 bp  
\* 21151 23505: contig of 2355 bp in length  
\* 23506 23605: gap of 100 bp  
\* 23606 34391: contig of 10786 bp in length  
\* 34392 34491: gap of 100 bp  
\* 34492 37041: contig of 2550 bp in length  
\* 37042 37141: gap of 100 bp  
\* 37142 40346: contig of 3205 bp in length  
\* 40347 40446: gap of 100 bp  
\* 40447 42517: contig of 2071 bp in length  
\* 42518 42617: gap of 100 bp  
\* 42618 44873: contig of 2256 bp in length  
\* 44874 44973: gap of 100 bp  
\* 44974 50196: contig of 5223 bp in length  
\* 50197 50296: gap of 100 bp  
\* 50297 52455: contig of 2159 bp in length  
\* 52456 52555: gap of 100 bp  
\* 52556 54810: contig of 2255 bp in length  
\* 54811 54910: gap of 100 bp  
\* 54911 60983: contig of 6073 bp in length  
\* 60984 61083: gap of 100 bp  
\* 61084 63186: contig of 2103 bp in length  
\* 63187 63286: gap of 100 bp  
\* 63287 66133: contig of 2847 bp in length  
\* 66134 66233: gap of 100 bp  
\* 66234 68778: contig of 2545 bp in length  
\* 68779 68878: gap of 100 bp  
\* 68879 71439: contig of 2561 bp in length  
\* 71440 71539: gap of 100 bp  
\* 71540 74946: contig of 3407 bp in length  
\* 74947 75046: gap of 100 bp  
\* 75047 77413: contig of 2367 bp in length  
\* 77414 77513: gap of 100 bp  
\* 77514 79615: contig of 2102 bp in length

\* 79616 79715: gap of 100 bp  
\* 79716 85282: contig of 5567 bp in length  
\* 85283 85382: gap of 100 bp  
\* 85383 92460: contig of 7078 bp in length  
\* 92461 92560: gap of 100 bp  
\* 92561 94871: contig of 2311 bp in length  
\* 94872 94971: gap of 100 bp  
\* 94972 98241: contig of 3270 bp in length  
\* 98242 98341: gap of 100 bp  
\* 98342 101349: contig of 3008 bp in length  
\* 101350 101449: gap of 100 bp  
\* 101450 104159: contig of 2710 bp in length  
\* 104160 104259: gap of 100 bp  
\* 104260 106320: contig of 2061 bp in length  
\* 106321 106420: gap of 100 bp  
\* 106421 109042: contig of 2622 bp in length  
\* 109043 109142: gap of 100 bp  
\* 109143 112269: contig of 3127 bp in length.  
\* Location/Qualifiers  
\* source  
\* 1..112269  
\* /organism="Mus musculus"  
\* /db\_xref="taxon:10090"  
\* /chromosome="2"  
\* /clone="RP23-256D19"  
\* /clone\_lib="RPC1-23"  
\* 1..2026  
\* /note="assembly\_fragment:00009"  
\* 2127..4515  
\* /note="assembly\_fragment:00017"  
\* 4516..7195  
\* /note="assembly\_fragment:00080"  
\* 7296..14681  
\* /note="assembly\_fragment:00136"  
\* 14782..21050  
\* /note="assembly\_fragment:00161"  
\* 21151..23505  
\* /note="assembly\_fragment:00194"  
\* 23606..34391  
\* /note="assembly\_fragment:00228"  
\* 34492..37041  
\* /note="assembly\_fragment:00259"  
\* 37142..40346  
\* /note="assembly\_fragment:00299"  
\* 40447..42517  
\* /note="assembly\_fragment:00315"  
\* 42618..44873  
\* /note="assembly\_fragment:00317"  
\* 44974..50196  
\* /note="assembly\_fragment:00365"  
\* 50297..52455  
\* /note="assembly\_fragment:00379"  
\* 52556..54810  
\* /note="assembly\_fragment:00418"  
\* 54911..60983  
\* /note="assembly\_fragment:00492"  
\* 61084..63186  
\* /note="assembly\_fragment:00529"  
\* 63287..66133  
\* /note="assembly\_fragment:00533"  
\* 66234..68778  
\* /note="assembly\_fragment:00599"  
\* 68879..71439  
\* /note="assembly\_fragment:00642"  
\* 71540..74946  
\* /note="assembly\_fragment:00690.0"  
\* 75047..77413  
\* /note="assembly\_fragment:00702"  
\* 77514..79615  
\* /note="assembly\_fragment:00733"  
\* 79716..85282  
\* /note="assembly\_fragment:00764"  
\* 85383..92460  
\* /note="assembly\_fragment:00784"



```
misc_feature /note="assembly_name:Contig33"
3024. 4722
misc_feature /note="assembly_name:Contig34"
4823. 6381
misc_feature /note="assembly_name:Contig36"
6482. 7910
misc_feature /note="assembly_name:Contig40"
8011. 9503
misc_feature /note="assembly_name:Contig42"
9604. 11058
misc_feature /note="assembly_name:Contig43"
11159. 13254
misc_feature /note="assembly_name:Contig44"
13355. 14882
misc_feature /note="assembly_name:Contig45"
14983. 16254
misc_feature /note="assembly_name:Contig46"
16355. 18120
misc_feature /note="assembly_name:Contig47"
18221. 19610
misc_feature /note="assembly_name:Contig48"
19711. 21191
misc_feature /note="assembly_name:Contig49"
21292. 23220
misc_feature /note="assembly_name:Contig50"
23321. 28270
misc_feature /note="assembly_name:Contig51"
26371. 28950
misc_feature /note="assembly_name:Contig52
clone_end.T7
vector_side:right"
misc_feature 29051. 32298
misc_feature /note="assembly_name:Contig53"
32399. 36137
misc_feature /note="assembly_name:Contig54"
36238. 39727
misc_feature /note="assembly_name:Contig55"
39828. 44103
misc_feature /note="assembly_name:Contig56"
44204. 47399
misc_feature /note="assembly_name:Contig57"
47500. 54396
misc_feature /note="assembly_name:Contig58"
54497. 60615
misc_feature /note="assembly_name:Contig59"
60716. 67010
misc_feature /note="assembly_name:Contig60"
67111. 78898
misc_feature /note="assembly_name:Contig61"
76999. 84028
misc_feature /note="assembly_name:Contig62"
84129. 97932
misc_feature /note="assembly_name:Contig63"
98033. 130383
misc_feature /note="assembly_name:Contig64"
130484. 171342
misc_feature /note="assembly_name:Contig65"
171343. 225433
misc_feature /note="assembly_name:Contig66"
225433. 2910 others

BASE COUNT 66412 a 45098 c 45266 g 65747 t 2910 others
ORIGIN

Query Match 69.5%; Score 26.4; DB 2; Length 225433;
Best Local Similarity 96.4%; Pred. No. 9;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
RESULT 8
AL845492/c
```

```
LOCUS      AL845492                245916 bp    DNA    linear    HTG 17-AUG-2002
DEFINITION Mus musculus chromosome 2 clone RP23-222P7, *** SEQUENCING IN
            PROGRESS ***, 11 unordered pieces.
ACCESSION  AL845492
VERSION    AL845492.2  GI:22416280
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1. (bases 1 to 245916)
            Plumb, B.
            Direct Submission
            Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Aug 21, 2002 this sequence version replaced gi:22316222.
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: BM222P7
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 243174 bases at least Q40
            Consensus quality: 243936 bases at least Q30
            Consensus quality: 244330 bases at least Q20
            Insert size: 244916; sum-of-contigs
            Insert size: 441095; 9.1% error; agarose-fp
            Quality coverage: 6.99x in Q20 bases; sum-of-contigs Quality
            coverage: 7.14x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 23764: contig of 23764 bp in length
            * 23765 23864: gap of 100 bp
            * 23865 54981: contig of 31117 bp in length
            * 54982 55081: gap of 100 bp
            * 55082 91767: contig of 36686 bp in length
            * 91768 91867: gap of 100 bp
            * 91868 94696: contig of 2829 bp in length
            * 94697 94796: gap of 100 bp
            * 94797 140303: contig of 45507 bp in length
            * 140304 140403: gap of 100 bp
            * 140404 144122: contig of 3719 bp in length
            * 144123 144222: gap of 100 bp
            * 144223 150853: contig of 6631 bp in length
            * 150854 150953: gap of 100 bp
            * 150954 192533: contig of 41580 bp in length
            * 192534 192633: gap of 100 bp
            * 192634 196792: contig of 4159 bp in length
            * 196793 196892: gap of 100 bp
            * 196893 200197: contig of 3305 bp in length
            * 200198 200297: gap of 100 bp
            * 200298 245916: contig of 45619 bp in length.
            *
            Location/Qualifiers
            1. 245916
               /organism="Mus musculus"
               /db_xref="taxon:10090"
               /chromosome="2"
               /clone="RP23-222P7"
               /clone_lib="RPC1-23"
               /note="assembly_fragment:02484
            misc_feature 1. 23764
```



```

misc_feature      fragment_chain:1"
23855...54981
/notes="assembly fragment:00096
fragment_chain:1"
55082...91767
/notes="assembly fragment:01352
fragment_chain:1"
91858...94696
/notes="assembly fragment:02266
fragment_chain:1"
94797...140303
/notes="assembly fragment:00090
fragment_chain:1"
140404...141122
/notes="assembly fragment:01481
fragment_chain:1"
144233...150853
/notes="assembly fragment:00446
fragment_chain:1"
150954...192533
/notes="assembly fragment:02940
fragment_chain:1"
192634...196792
/notes="assembly fragment:02221
fragment_chain:1"
196893...200197
/notes="assembly fragment:01308
fragment_chain:2"
200298...245916
/notes="assembly fragment:02802
fragment_chain:2
clone_end:T7
vector_side:right"
BASE COUNT      71028 a 49815 c 49968 g 74059 t 1006 others
ORIGIN

```

```

Query Match      69.5%; Score 26.4; DB 2; Length 245916;
Best Local Similarity 96.4%; Pred.No. 9.8;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 TCAAGGTCAAAACACTAGGTCAAAGGTCA 38
Db 181912 TCAAGGTCAAAACACTAGGTCAAAGGTCA 181885

```

```

RESULT 9
AC107711
LOCUS            210768 bp DNA linear HTG 26-APR-2002
DEFINITION      Mus musculus clone RP23-291112, WORKING DRAFT SEQUENCE, 23 ordered
pieces.
AC107711
AC107711.3 GI:20330925
KEYWORDS         HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE          Mus musculus.
ORGANISM        Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 210768)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-291112
Unpublished
REFERENCE
2 (bases 1 to 210768)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chararo,B.,
Chapel,I.Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczyk,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,P., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppach,R.,
Seman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center code: WfBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19049
Center clone name: 291_112
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 204366 bases at least Q40
Consensus quality: 206843 bases at least Q30
Consensus quality: 207569 bases at least Q20
Insert size: 208000; agarose-fp
Quality coverage: 6.9 in Q20 bases; sum-of-contigs
Quality coverage: 6.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

```

McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,P., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 210768)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,P.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppach,R.,
Seman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center code: WfBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19049
Center clone name: 291_112
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 204366 bases at least Q40
Consensus quality: 206843 bases at least Q30
Consensus quality: 207569 bases at least Q20
Insert size: 208000; agarose-fp
Quality coverage: 6.9 in Q20 bases; sum-of-contigs
Quality coverage: 6.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

```

* 1 155: contig of 155 bp in length
* 156 255: gap of 100 bp
* 256 946: contig of 691 bp in length
* 947 1046: gap of 100 bp
* 1047 1677: contig of 631 bp in length
* 1678 1777: gap of 100 bp
* 1778 2851: contig of 1074 bp in length
* 2852 2951: gap of 100 bp
* 2952 4519: contig of 1568 bp in length
* 4520 4619: gap of 100 bp
* 4620 5830: contig of 1211 bp in length
* 5831 5930: gap of 100 bp
* 5931 7289: contig of 1359 bp in length
* 7290 7389: gap of 100 bp
* 7390 8943: contig of 1554 bp in length
* 8944 9043: gap of 100 bp
* 9044 10541: contig of 1498 bp in length
* 10542 10641: gap of 100 bp
* 10642 12527: contig of 1886 bp in length
* 12528 12627: gap of 100 bp
* 12628 15560: contig of 2933 bp in length
* 15561 15660: gap of 100 bp
* 15661 19236: contig of 3576 bp in length
* 19237 19336: gap of 100 bp
* 19337 22891: contig of 3555 bp in length
* 22892 22991: gap of 100 bp
* 22992 28481: contig of 5490 bp in length
* 28482 28581: gap of 100 bp
* 28582 39209: contig of 10628 bp in length
* 39210 39309: gap of 100 bp
* 39310 48823: contig of 9514 bp in length
* 48824 48923: gap of 100 bp
* 48924 61869: contig of 12946 bp in length
* 61870 61969: gap of 100 bp
* 61970 76596: contig of 14727 bp in length
* 76597 76796: gap of 100 bp
* 76797 99351: contig of 23155 bp in length
* 99352 100051: gap of 100 bp
* 100052 122918: contig of 22867 bp in length
* 122919 123018: gap of 100 bp
* 123019 152004: contig of 28986 bp in length
* 152005 152104: gap of 100 bp
* 152105 183845: contig of 31741 bp in length
* 183846 183945: gap of 100 bp
* 183946 210768: contig of 26823 bp in length.
FEATURES
    source
        1..210768
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-291112"
            /clone_lib="RPCI-23 Female Mouse BAC"
        1..155
            /note="assembly_fragment"
            clone end:SP6
            vector side:left
            256..946
                /note="assembly_fragment"
            1047..1677
                /note="assembly_fragment"
            1778..2851
                /note="assembly_fragment"
            2952..4519
                /note="assembly_fragment"
            4620..5830
                /note="assembly_fragment"
            5931..7289
                /note="assembly_fragment"
            7390..8943
                /note="assembly_fragment"
            9044..10541
                /note="assembly_fragment"
            10642..12527
                /note="assembly_fragment"
            12628..15560
                /note="assembly_fragment"
            15661..19236
                /note="assembly_fragment"
            19237..22891
                /note="assembly_fragment"
            22892..28481
                /note="assembly_fragment"
            28482..39209
                /note="assembly_fragment"
            39210..48823
                /note="assembly_fragment"
            48824..48923
                /note="assembly_fragment"
            48924..61869
                /note="assembly_fragment"
            61870..76596
                /note="assembly_fragment"
            76597..76796
                /note="assembly_fragment"
            76797..99351
                /note="assembly_fragment"
            99352..100051
                /note="assembly_fragment"
            100052..122918
                /note="assembly_fragment"
            122919..123018
                /note="assembly_fragment"
            123019..152004
                /note="assembly_fragment"
            152005..183845
                /note="assembly_fragment"
            183846..183945
                /note="assembly_fragment"
            183946..210768
                /note="assembly_fragment"
            clone end:T7
            vector_side:right
    misc_feature
        12628..15560
            /note="assembly_fragment"
        15661..19236
            /note="assembly_fragment"
        19337..22891
            /note="assembly_fragment"
        22992..28481
            /note="assembly_fragment"
        28582..39209
            /note="assembly_fragment"
        39310..48823
            /note="assembly_fragment"
        48924..61869
            /note="assembly_fragment"
        61970..76596
            /note="assembly_fragment"
        76797..99351
            /note="assembly_fragment"
        100052..122918
            /note="assembly_fragment"
        123019..152004
            /note="assembly_fragment"
        152105..183845
            /note="assembly_fragment"
        183946..210768
            /note="assembly_fragment"
        clone end:T7
        vector_side:right
    Query Match
        68.4%; Score 26; DB 2; Length 210768;
    Best Local Similarity
        85.3%; Pred. No. 13;
    Matches
        29; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
    QY
        1 CAAAAGTGGTCAAGGTCATACTGGTCAAG 34
    Db
        32218 CAAAAGTGGTCAAGGTCATACTGGTCAAG 32251
    RESULT 10
    AX251575
    LOCUS
        AX251575 41 bp DNA linear PAT 05-OCT-2001
    DEFINITION
        Sequence 3 from Patent WO0168845.
    ACCESSION
        AX251575
    VERSION
        AX251575.1 GI:15984998
    KEYWORDS
        synthetic construct.
        synthetic construct
        artificial sequences.
        1 (bases 1 to 41)
    AUTHORS
        Massaad, C., Berenbaum, F., Olivier, J.L., Salvat, C. and Bereziat, G.
    TITLE
        Inflammation-inducible hybrid promoters, vectors containing same
        and uses thereof
    JOURNAL
        Patent: WO 0168845-A 3 20-SEP-2001;
        Aventis Pharma S.A. (FR)
    FEATURES
        Location/Qualifiers
            1..41
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="element PPPE"
            18 a
                /note="element PPPE"
            7 t
    BASE COUNT
        18 a 8 c 8 g 7 t
    ORIGIN
    Query Match
        65.8%; Score 25; DB 6; Length 41;
    Best Local Similarity
        92.7%; Pred. No. 1.5e+02;
    Matches
        38; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
    QY
        1 CAAAAGTGGTCAAGGTCATACTGGTCAAGTCA 38
    Db
        1 CAAAAGTGGTCAAGGTCATACTGGTCAAGTCA 41
    RESULT 11
    AX251578

```

LOCUS AX251578 332 bp DNA linear PAT 05-OCT-2001  
 DEFINITION Sequence 6 from Patent WO0168845.  
 ACCESSION AX251578  
 VERSION AX251578.1 GI:15985001  
 SOURCE synthetic construct.  
 ORGANISM synthetic construct.  
 REFERENCE 1 (bases 1 to 332)  
 AUTHORS Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Berezziat,G.  
 TITLE Inflammation-inducible hybrid promoters, vectors containing same and uses thereof  
 JOURNAL Patent: WO 0168845-A 6 20-SEP-2001;  
 Aventis Pharma S.A. (FR)  
 FEATURES  
 source  
 1. .332  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="promoteur hybride PPRE/PLA2s"  
 BASE COUNT 96 a 91 c 82 g 63 t  
 ORIGIN  
 Query Match 65.8%; Score 25; DB 6; Length 332;  
 Best Local Similarity 92.7%; Pred. No. 1e+02;  
 Matches 38; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
 QY 1 CAAACTAGGTCAAAGC---TCAAACTAGGTCAAAGGTCA 38  
 Db 13 CAAACTAGGTCAAAGTCATCAAACTAGGTCAAAGGTCA 53  
 RESULT 12  
 AX063386  
 LOCUS AX063386 59 bp DNA linear PAT 24-JAN-2001  
 DEFINITION Sequence 6 from Patent WO0078986.  
 ACCESSION AX063386  
 VERSION AX063386.1 GI:12541176  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 69)  
 AUTHORS Dartell,R., Crouzet,J., Staels,B. and Mahfoudi,A.  
 TITLE Regulation system of expression using nuclear ppar receptors  
 JOURNAL Patent: WO 0078986-A 6 28-DEC-2000;  
 Aventis Pharma S.A. (FR)  
 FEATURES  
 source  
 1. .69  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 27 a 14 c 17 g 11 t  
 ORIGIN  
 Query Match 63.2%; Score 24; DB 6; Length 69;  
 Best Local Similarity 90.5%; Pred. No. 3.1e+02;  
 Matches 38; Conservative 0; Mismatches 0; Indels 4; Gaps 1;  
 QY 1 CAAACTAGGTCAAAGTC---AAACTAGGTCAAAGGTCA 38  
 Db 17 CAAACTAGGTCAAAGGTCACGGAACAACTAGGTCAAAGGTCA 58  
 RESULT 13  
 AC131487  
 LOCUS AC131487 101913 bp DNA linear HTG 23-AUG-2002  
 DEFINITION Lytechinus variegatus clone Lv18J3, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 4 ordered pieces.  
 ACCESSION AC131487  
 VERSION AC131487.1 GI:22450539  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_FULLTOP.  
 SOURCE green urchin.  
 ORGANISM Lytechinus variegatus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Euechinoidea; Echinacea; Temnopleuroidea;  
 Toxopneustidae; Lytechinus.  
 REFERENCE 1 (bases 1 to 101913)  
 AUTHORS Davidson,E.H., Rast,J.P., Oliveri,P., Ransick,A., Calestani,C.,  
 Yuh,C.-H., Minokawa,T., Amore,G., Hinman,V., Arenas-Mena,C.,  
 Otm,C., Brown,C.Titus., Livi,C.B., Lee,P.Y., Revilla,R.,  
 Schilstra,M.J., Clarke,P.J.C., Rust,A.G., Pan,Z., Arnone,M.I.,  
 Rowen,J., Cameron,R.Andrew., McClay,D.R., Hood,L. and Bolouri,H.  
 TITLE A provisional regulatory gene network for specification of  
 endomesoderm in the sea urchin embryo  
 JOURNAL Dev. Biol. 246 (1), 162-190 (2002)  
 MEDLINE 20204154  
 REFERENCE 2 (bases 1 to 101913)  
 AUTHORS Rowen,L., Cameron,R.A. and Davidson,E.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-AUG-2002) Multimegabase Sequencing Center, Institute  
 for Systems Biology, 1441 N. 34th Street, Seattle, WA 98103, USA  
 COMMENT  
 ----- Genome Center  
 Center: Multimegabase Sequencing Center  
 Center code: ISBMS  
 Web site: http://www.systemsbio.org  
 Contact: leerowen@systemsbio.org  
 Drafting center: Institute for Systems Biology (ISBMS)  
 ----- Summary Statistics  
 Sequencing vector: pUC18; L08752  
 Chemistry: Dye-terminator Big Dye; 90% of reads  
 Chemistry: Dye-primer Big Dye; 10% of reads  
 Assembly program: Phrap; version 0.990399  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 74199: contig of 74199 bp in length  
 \* 74200 74299: gap of unknown length  
 \* 74300 77966: contig of 3667 bp in length  
 \* 77967 78066: gap of unknown length  
 \* 78067 78459: contig of 393 bp in length  
 \* 78460 78559: gap of unknown length  
 \* 78560 101913: contig of 23354 bp in length.  
 FEATURES  
 Location/Qualifiers  
 1..101913  
 /organism="Lytechinus variegatus"  
 /db\_xref="taxon:7654"  
 /clone="Lv18J3"  
 /clone\_lib="Caltech Lytechinus variegatus sperm genomic  
 BAC library A"  
 /note="This library is described in Davidson et al.,  
 Developmental Biology 246, 162-190 (2002)."  
 BASE COUNT 31787 a 19032 c 18538 g 32255 t 301 others  
 ORIGIN  
 Query Match 63.2%; Score 24; DB 2; Length 101913;  
 Best Local Similarity 84.4%; Pred. No. 74;  
 Matches 27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 7 TAGGTCAAAGGTCAAACCTAGGTCAAAGGTCA 38  
 Db 69984 TTGGTCAAAGGTCAATACCGGTCAAAGGTCA 70015  
 RESULT 14  
 AC126043  
 LOCUS AC126043 153172 bp DNA linear HTG 02-JUL-2002  
 DEFINITION Mus musculus chromosome UNK clone RP24-356N5, WORKING DRAFT  
 SEQUENCE, 5 unordered pieces.  
 ACCESSION AC126043

VERSION AC126043.1 GI:21672241  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 153172)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE The sequence of Mus musculus clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 153172)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Project Information -----  
Center project name: M\_EB0356N05

----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Chemistry: Dye-primer ET; 0% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 150681 bases at least Q40  
Consensus quality: 150794 bases at least Q30  
Consensus quality: 150878 bases at least Q20  
Insert size: 140000; agarose-fp  
Insert size: 152772; sum-of-contigs  
Quality coverage: 14.63 in Q20 bases; agarose-fp  
Quality coverage: 12.60 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 5757: contig of 5757 bp in length  
\* 5758 5857: gap of unknown length  
\* 5858 21868: contig of 16011 bp in length  
\* 21869 21969: gap of unknown length  
\* 21969 53154: contig of 31185 bp in length  
\* 53154 53254: gap of unknown length  
\* 53254 97648: contig of 44395 bp in length  
\* 97648 97749: gap of unknown length  
\* 97749 153172: contig of 55424 bp in length.

FEATURES  
Location/Qualifiers  
1..153172  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clone="RP24-356N5"

misc\_feature 1..5757  
/note="assembly\_name:Contig5"  
misc\_feature 5858..21868  
/note="assembly\_name:Contig6"  
misc\_feature 21969..53153  
/note="assembly\_name:Contig7"  
misc\_feature 53254..97648  
/note="assembly\_name:Contig8"  
misc\_feature 97749..153172  
/note="assembly\_name:Contig9"  
BASE COUNT 45268 a 31701 c 31183 g 44616 t 404 others  
ORIGIN

Query Match 62.6%; Score 23.8; DB 2; Length 153172;  
Best Local Similarity 80.0%; Pred. No. 80;  
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 4 AACTAGGTCAAAGGTCACAACTAGGTCAAAGGTCA 38  
|||||  
DB 136601 AACTAGGTCAAAGGTCACAACTAGGTC 136635  
|||||  
RESULT 15  
HSU63289 2113 bp mRNA linear PRI 12-DEC-1996  
LOCUS Human RNA-binding protein CUG-BP/hNab50 (NAB50) mRNA, complete cds.  
DEFINITION U63289  
ACCESSION U63289.1 GI:1518801  
VERSION  
KEYWORDS  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2113)  
AUTHORS Timchenko,L.T., Miller,J.W., Timchenko,N.A., DeVore,D.R.,  
Datar,K.V., Lin,L., Roberts,R., Caskey,C.T. and Swanson,M.S.  
TITLE Identification of a (CUG)n triplet repeat RNA-binding protein and  
its expression in myotonic dystrophy  
JOURNAL Nucleic Acids Res. 24 (22), 4407-4414 (1996)  
MEDLINE 97105883  
PubMed 8948831  
REFERENCE 2 (bases 1 to 2113)  
AUTHORS Timchenko,L.T., Miller,J.W., Timchenko,N.A., DeVore,D.R.,  
Datar,K.V., Lin,L., Roberts,R., Caskey,C.T. and Swanson,M.S.  
TITLE Direct Submission  
JOURNAL Submitted (08-JUL-1996) Molecular Genetics and Microbiology,  
University of Florida, 1600 SW Archer Road, Gainesville, FL  
32610-0266, USA

FEATURES  
Location/Qualifiers  
1..2113  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="osteosarcoma cell line"  
1..2113  
/gene="NAB50"  
138..1586  
/gene="NAB50"  
/function="binds to (CUG)n triplet repeats"  
/note="RNA-binding protein"  
/codon\_start=1  
/product="CUG-BP/hNab50"  
/protein\_id="AAC50895.1"  
/db\_xref="GI:1518802"

translation="MNGTLDHPDQDLDAIKMFVQVPRTWSEKDLRELFRQYGAAYE  
INVLDRSQNPFSKGCCTVFYTRKAALEAQNALHNKVLFGMHHP:IQMKPADSEKN  
NAVEDRLFTGMSKKCTENDIRVNFSSFGQTEICRLRPGDGLSRGCAFYFTFRAM  
AQTAKHQAQTMGCSPPMVVFADTKQKQKMAQQLQQQMQQISAAASVMTNLA  
GLNTLQVALLQQTASSGNLNTLSLHPMGLNAMLONLALAAASAAQNTPSG  
TNALTSSSPLSVLSSGSSSSSSSNVPIASIGALOTLAGATAG:NVGSLAGMAA  
LNGGLGSGLSNGTSGTMEALTAQSGQQVAAALPTLYNQNLITQOSIGASQKE  
GPEGANLFYHLFPQFQDQLQMFPFGVNVSAKFVFDKQTNUSKCFGFSYDNPVS  
AAQAISMGNGFQIMKRLKVLQRKSNPSKY"  
BASE COUNT 523 a 518 c 523 g 549 t  
ORIGIN

Query Match 62.1%; Score 23.6; DB 9; Length 2113;  
Best Local Similarity 76.3%; Pred. No. 2.2e+02;  
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 1 CAAACTAGGTCAAAGGTCACAACTAGGTCAAAGGTCA 38  
|||||  
DB 1764 CACATCAGCATCAAGGTCACACACAGGTCAAGGTCA 1727  
|||||

Search completed: February 19, 2003, 22:55:50



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:26:38 ; Search time 23.0954 Seconds  
(without alignments)  
3705.323 Million cell updates/sec

Title: US-09-808-388-2  
Perfect score: 38  
Sequence: 1 caaaactaggccaaggtcaaaactaggtcaaggtca 38

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	38	22	AAI64304
2	25	65.8	41	22	AAI64305
3	25	65.8	332	22	AAI64308
4	24	63.2	69	23	ABL58060
5	23.6	62.1	2886	24	ABO54898
6	23.6	62.1	12961	22	ABN20012
7	23.2	61.1	64	23	ABL58061
8	22.6	59.5	731	13	AAQ21036
9	22	57.9	413	22	AAI18762

10	22	57.9	5957	10	AAI90002	Nucleotide Sequenc
11	22	57.9	5957	10	AAI90002	Nucleotide Sequenc
12	21.6	56.8	2495	24	ABL53738	Human cell factor
13	21.6	56.8	2647	22	AAH17795	Human cDNA sequenc
14	21.6	56.8	2939	21	AAI76931	Human ORFX ORF2486
15	21.4	56.3	14654	22	AAI28620	Genomic sequence #
16	21.2	55.8	3192	22	AAH54618	S. epidermidis gen
17	21	55.3	399	22	AAI88608	Human digestive by
18	21	55.3	11874	22	AAI88608	Human immune/haema
19	21	55.3	75899	24	AAK85723	Human immune/haema
20	21	55.3	75899	24	AAK85723	Human immune/haema
21	20.8	54.7	1235	21	AAI25342	Human genomic DNA
22	20.8	54.7	1235	21	AAI25342	Human secreted pro
23	20.8	54.7	1698	21	AAI32964	Arabidopsis thalia
24	20.6	54.2	425	24	AAI35112	Human EST (expres
25	20.6	54.2	10953	18	AAI74373	Staphylococcus aur
26	20.4	53.7	828	22	AAI93963	Human neuroblastom
27	20.4	53.7	954	23	AAI83540	DNA encoding novel
28	20.4	53.7	3539	21	AAI33643	TT virus (TTV-JA10
29	20.4	53.7	3853	21	AAI33638	TT virus (TTV-JA20
30	20.4	53.7	4796	21	AAI28840	Human T cell induc
31	20.4	53.7	4797	22	AAI14876	Human TIF genomic
32	20.4	53.7	4797	24	AAI30646	Human T cell deriv
33	20.4	53.7	4797	24	AAI27151	Drosophila melanog
34	20.4	53.7	37996	23	AAI07876	PPAR response elem
35	20.2	53.2	52	22	AAI64306	Corn tassal-derive
36	20.2	53.2	94	24	ABL71021	Human foetal liver
37	20.2	53.2	480	22	ABA57934	Human brain expres
38	20.2	53.2	480	22	AAK06007	Human bone marrow
39	20.2	53.2	480	22	AAK1651	Probe #6214 used t
40	20.2	53.2	480	22	AAI37528	Human genome-deriv
41	20.2	53.2	480	24	ABS06456	Streptococcus pneu
42	20.2	53.2	541	18	AAI30736	Human liver cell s
43	20.2	53.2	1705	22	AAI57482	Streptococcus pneu
44	20.2	53.2	11443	19	AAI52182	Drosophila melanog
45	20.2	53.2	32328	23	ABL09994	

ALIGNMENTS

RESULT 1  
AAI64304  
ID AAI64304 standard; DNA; 38 BP.  
XX  
AC AAI64304;  
DT 15-NOV-2001 (first entry)  
DE PPAR response element (DR1)2 17.  
XX  
XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
XX cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;  
XX peroxisome proliferator activated receptor;  
XX secreted non-pancreatic phospholipase A2; ss.  
XX Synthetic.  
XX OS  
XX WO200168845-A2.  
XX  
XX 20-SEP-2001.  
XX  
XX 14-MAR-2001; 2001WO-FR00759.  
XX  
XX 14-MAR-2000; 2000FR-0003262.  
XX 13-APR-2000; 2000US-0196959.  
(AVET ) AVENTIS PHARMA SA.  
Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;  
WPI; 2001-582451/65.



RESULT 4  
 ABQ54896  
 ID ABL58060 standard; DNA; 69 BP.  
 XX  
 AC ABL58060;  
 XX  
 DT 22-JUL-2002 (first entry)  
 DE Human PPAR response element consensus DR PCR primer 1RDM69.  
 XX  
 KW PPAR response element; PPAR; vaccine; gene therapy; human;  
 KW peroxisome proliferator-activated receptor; promoter; PCR; primer; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200078986-A1.  
 XX  
 PD 28-DEC-2000.  
 XX  
 PF 22-JUN-2000; 2000WO-FR01744.  
 XX  
 PR 22-JUN-1999; 99FR-0007957.  
 PR 20-AUG-1999; 99US-0149721.  
 XX  
 PA (AVET ) AVENTIS PHARMA SA.  
 PI Darteil R, Crouzet J, Staels B, Mahfoudi A;  
 XX  
 DR WPI; 2001-091574/10.  
 XX  
 PT Composition providing inducible expression of a nucleic acid, useful in  
 PT gene therapy, uses minimal promoter with peroxisome  
 PT proliferator-activated receptor response elements -  
 XX  
 PS Example 1; Page 29; 94pp; French.  
 XX  
 CC The present invention relates to a composition (A) comprising a component  
 CC (A1) containing a nucleic acid (I) controlled by an inducible promoter  
 CC that consists of a PPAR (peroxisome proliferator-activated receptor)  
 CC response element (ABL58055) and a minimal promoter; and/or a component  
 CC (A2) comprising a nucleic acid encoding a PPAR under control of a  
 CC transcriptional promoter. (A) and vectors containing (A1) and (A2), are  
 CC used to express (I) in cells for expression of transgenic (I) for  
 CC experimental, clinical, therapeutic or diagnostic purposes. (I) encodes  
 CC an agriculturally useful, therapeutic, vaccinating or marker protein and  
 CC is most especially expressed in human muscle cells. Cells containing (A),  
 CC or the vectors, are used to identify PPAR ligands or to produce  
 CC transgenic animals for preclinical studies, analysis of bioavailability,  
 CC labelling etc. The present sequence is PCR primer, for human PPAR  
 CC response element consensus DR, which was used in an example from the  
 CC invention.  
 XX  
 SQ Sequence 69 BP; 27 A; 14 C; 17 G; 11 T; 0 other;  
 Query Match 63.2%; Score 24; DB 23; Length 69;  
 Best Local Similarity 90.5%; Pred. No. 3.7;  
 Matches 38; Conservative 0; Mismatches 0; Indels 4; Gaps 1;  
 QY 1 CAAACTAGGTCAAAGGTC-----AAACTAGGTCAAAGGTC 38  
 DB 17 CAAACTAGGTCAAAGGTCACGGAAACTAGGTCAAAGGTC 58  
 RESULT 5  
 ABQ54898/c  
 ID ABQ54898 standard; cDNA; 2886 BP.  
 XX  
 AC ABQ54898;  
 XX  
 DT 22-AUG-2002 (first entry)  
 XX

DE Human ovarian antigen HAZAA59 cDNA, SEQ ID NO:778.  
 XX  
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200677-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 07-JUN-2001; 2001WO-US18569.  
 XX  
 PR 07-JUN-2000; 2000US-209467P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 DR WPI; 2002-147878/19.  
 DR P-PSDB; ABP41821.  
 XX  
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
 PT ovarian cancer), immune disorders, cardiovascular disorders and  
 PT neurological diseases -  
 XX  
 PS Claim 1; SEQ ID No 778; 2322pp; English.  
 XX  
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ58305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention.  
 CC  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2886 BP; 680 A; 615 C; 619 G; 971 T; 1 other;  
 Query Match 62.1%; Score 23.6; DB 24; Length 2886;  
 Best Local Similarity 76.3%; Pred. No. 9.4;  
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;



QY 1 CAAAAGTGGTCAAGGTCAAAAGTGGTCAAGGTC 38  
 Db 926 CACATCAGCATCAAGGTCACACACAGGTCAGGTCAGGTC 889

RESULT 6  
 ABA20012/c  
 ID ABA20012 standard; DNA; 12961 BP.

XX AC ABA20012;

DT 23-JAN-2002 (first entry)

XX DE Human nervous system related polynucleotide SEQ ID NO 12343.

XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antitickling; antianemic; antiarthritic; cancer;  
 KW antineumatic; hepatotropic; cerebroprotective; antinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX OS Homo sapiens.

XX PN WO200159063-A2.

XX PD 16-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01334.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 14-JUL-2000; 2000US-0217486.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

05-SEP-2000; 2000US-0229513.  
 06-SEP-2000; 2000US-0230437.  
 06-SEP-2000; 2000US-0230438.  
 08-SEP-2000; 2000US-0231242.  
 08-SEP-2000; 2000US-0231243.  
 08-SEP-2000; 2000US-0231244.  
 08-SEP-2000; 2000US-0231413.  
 08-SEP-2000; 2000US-0231414.  
 08-SEP-2000; 2000US-0232080.  
 08-SEP-2000; 2000US-0232081.  
 12-SEP-2000; 2000US-0231968.  
 14-SEP-2000; 2000US-0232397.  
 14-SEP-2000; 2000US-0232398.  
 14-SEP-2000; 2000US-0232399.  
 14-SEP-2000; 2000US-0232400.  
 14-SEP-2000; 2000US-0232401.  
 14-SEP-2000; 2000US-0233063.  
 14-SEP-2000; 2000US-0233064.  
 14-SEP-2000; 2000US-0233065.  
 21-SEP-2000; 2000US-0234223.  
 21-SEP-2000; 2000US-0234274.  
 25-SEP-2000; 2000US-0234997.  
 25-SEP-2000; 2000US-0234998.  
 26-SEP-2000; 2000US-0235484.  
 27-SEP-2000; 2000US-0235834.  
 27-SEP-2000; 2000US-0235836.  
 29-SEP-2000; 2000US-0236327.  
 29-SEP-2000; 2000US-0236367.  
 29-SEP-2000; 2000US-0236368.  
 29-SEP-2000; 2000US-0236369.  
 29-SEP-2000; 2000US-0236370.  
 02-OCT-2000; 2000US-0236802.  
 02-OCT-2000; 2000US-0237037.  
 02-OCT-2000; 2000US-0237038.  
 02-OCT-2000; 2000US-0237039.  
 02-OCT-2000; 2000US-0237040.  
 13-OCT-2000; 2000US-0239935.  
 13-OCT-2000; 2000US-0239937.  
 20-OCT-2000; 2000US-0240960.  
 20-OCT-2000; 2000US-0241785.  
 20-OCT-2000; 2000US-0241786.  
 20-OCT-2000; 2000US-0241787.  
 20-OCT-2000; 2000US-0241808.  
 20-OCT-2000; 2000US-0241809.  
 20-OCT-2000; 2000US-0241826.  
 20-OCT-2000; 2000US-0242221.  
 01-NOV-2000; 2000US-0244617.  
 08-NOV-2000; 2000US-0246474.  
 08-NOV-2000; 2000US-0246475.  
 08-NOV-2000; 2000US-0246476.  
 08-NOV-2000; 2000US-0246477.  
 08-NOV-2000; 2000US-0246478.  
 08-NOV-2000; 2000US-0246523.  
 08-NOV-2000; 2000US-0246524.  
 08-NOV-2000; 2000US-0246525.  
 08-NOV-2000; 2000US-0246526.  
 08-NOV-2000; 2000US-0246527.  
 08-NOV-2000; 2000US-0246528.  
 08-NOV-2000; 2000US-0246532.  
 08-NOV-2000; 2000US-0246609.  
 08-NOV-2000; 2000US-0246610.  
 08-NOV-2000; 2000US-0246611.  
 08-NOV-2000; 2000US-0246613.  
 17-NOV-2000; 2000US-0249207.  
 17-NOV-2000; 2000US-0249208.  
 17-NOV-2000; 2000US-0249209.  
 17-NOV-2000; 2000US-0249210.  
 17-NOV-2000; 2000US-0249211.  
 17-NOV-2000; 2000US-0249212.  
 17-NOV-2000; 2000US-0249213.  
 17-NOV-2000; 2000US-0249214.  
 17-NOV-2000; 2000US-0249215.  
 17-NOV-2000; 2000US-0249216.



PF 24-APR-1990; 90JP-0106354.

XX 24-APR-1990; 90JP-0106354.

PR (TAKA-) TAKARA SHUZO KK.

PA WPI; 1992-060511/08.

XX Detecting mycoplasma - by detecting and amplifying DNA sequence

PT coding for mycoplasma and RNA

XX Example 2; Fig 9; 13pp; Japanese.

XX This region of M.hypopneumoniae-specific DNA was amplified by PCR.

CC See also AAQ21028-Q21051.

XX Sequence 731 BP; 252 A; 94 C; 103 G; 259 T; 23 other;

Query Match 59.5%; Score 22.6; DB 13; Length 731;

Best Local Similarity 75.7%; Pred. No. 18;

Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CAAAACTAGGTCAAAGGTCAAAAGTCAAGTCAAGGTC 37

DB 420 CAAAACTAGGTCATATAGCCAAAAGTCAAGTCAAGTCAAGTCA 384

RESULT 9

AAAL18762/C

ID AAL18762 standard; cDNA; 413 BP.

XX AAL18762;

XX 07-DEC-2001 (first entry)

DT Human breast cancer expressed polynucleotide 11219.

DE Human; breast cancer; cell marker; cytostatic; ss.

KW Homo sapiens.

XX WO200151628-A2.

XX 19-JUL-2001.

XX 10-JAN-2001; 2001WO-US00798.

XX 14-JAN-2000; 2000US-0176077.

XX 14-MAR-2000; 2000US-0189167.

XX 24-MAR-2000; 2000US-0192099.

XX 29-MAR-2000; 2000US-0193480.

XX 15-MAY-2000; 2000US-0205230.

XX 09-JUN-2000; 2000US-0211315.

XX 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer -

PS Claim 1; Page 2000; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides

CC (AAL07544-AAL26789) and methods of assessing whether a patient is

CC afflicted with breast cancer by examining the correlation between the

CC expression of certain markers and the cancerous state of breast cells.

CC The polynucleotides and encoded polypeptides are potential markers for

CC detecting, diagnosing, monitoring, characterising treating and

CC potentially preventing breast cancer. The polynucleotides and encoded

CC polypeptides are also useful for isolating compounds with cytostatic

CC activity.

XX Sequence 413 BP; 73 A; 124 C; 88 G; 128 T; 0 other;

SQ Query Match 57.9%; Score 22; DB 22; Length 413;

Best Local Similarity 73.7%; Pred. No. 28;

Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CAAAACTAGGTCAAAGGTCAAAAGTCAAGTCAAGGTC 38

DB 44 CCAAACTAGGTCAAAGGTCATATAGGTCATATAGGTC 7

RESULT 10

AAAN90002

ID AAN90002 standard; DNA; 5957 BP.

XX AAN90002;

XX 01-NOV-1989 (first entry)

DT Nucleotide Sequence of Denguevirus J.

DE Denguevirus J; insecticide; Spodoptera.

KW Denonucleosis virus J.

XX BP319418-A.

XX 07-JUN-1989.

XX 01-DEC-1987; 87EP-0403030.

XX 03-DEC-1987; 87EP-0403030.

XX (ROUS ) ROUSSEL UCLAF.

XX Gervais M, Bergoin M, Jourdan M, Jousset FX;

XX WPI; 1989-167414/23.

XX New recombinant plasmid contg. denguevirus DNA

PT - can be replicated in E. coli and useful as

PT biological insecticide.

XX Claim 4; page 7; 18pp; French.

XX Nucleotide sequence of Denguevirus J (originating from

CC Junonia) which causes denguevirus in susceptible insects.

CC Used in plasmids for the biological control of, eg Spodoptera,

CC including strains resistant to chemical insecticides.

XX Sequence 5957 BP; 1867 A; 1153 C; 1047 G; 1888 T; 2 other;

SQ Query Match 57.9%; Score 22; DB 10; Length 5957;

Best Local Similarity 73.7%; Pred. No. 43;

Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CAAAACTAGGTCAAAGGTCAAAAGTCAAGTCAAGGTC 38

DB 5953 CAGAACTAGGTCATATAGGTCATATAGGTCATATAGGTC 5990

RESULT 11

AAAN90002/c

ID AAN90002 standard; DNA; 5957 BP.

XX AAN90002;

XX 01-NOV-1989 (first entry)

DT Nucleotide Sequence of Denguevirus J.

XX Denguevirus J; insecticide; Spodoptera.



complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AA03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AB02446 to AAH13632 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Query Match 2647 BP; 808 A; 518 C; 622 G; 699 T; 0 other;  
 Query Match 56.8%; Score 21.6; DB 22; Length 2647;  
 Best Local Similarity 75.0%; Pred. No. 54;  
 Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

3 AACCTAGTCAAGTCAAACTAGGTCAGGTCGA 38

949 AACCAAGTCAAGGCCATGCTATGCTAGGTCGA 984

## RESULT 14

AAC76931  
 ID AAC76931 standard; cDNA; 2939 BP.

AC AAC76931;

08-FEB-2001 (first entry)

Human ORFX ORF2486 polynucleotide sequence SEQ ID NO:4971.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antiparotatic; antiparkinsonian; nontropic; neuroprotective; anticonvulsant; osteopathic; aniaarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antinflammatory; antiviral; antifungal; antitumor; antirheumatic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 95US-0127607.

02-APR-1999; 95US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach M;

WPI; 2000-602362/57.

P-PSDB; AAB42722.

Novel nucleic acids and peptides derived from open reading frame X,

useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -  
 Claim 5; Page 4141-4143; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnary; antiparotatic; antiparkinsonian; nontropic; neuroprotective; osteopathic; anticonvulsant; aniaarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating CC pathologic conditions associated with an ORFX-associated disorder. The CC nucleic acids can be used to express ORFX proteins in gene therapy CC vectors. The proteins and nucleic acids may be used to treat cancers, CC proliferative disorders, neurodegenerative disorders, osteoarthritis, CC graft vs host disease, cardiovascular disease, diabetes mellitus, lupus CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, CC bacterial or fungal infection, malaria, autoimmune disorders, asthma, CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, CC nocturnal haemoglobinuria, antinflammatory disease; to enhance CC coagulation; to inhibit thrombosis; and as a contraceptive.

Query Match 2939 BP; 904 A; 579 C; 687 G; 769 T; 0 other;

Query Match 56.8%; Score 21.6; DB 21; Length 2939;

Best Local Similarity 75.0%; Pred. No. 55;

Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

3 AACCTAGTCAAGTCAAACTAGGTCAGGTCGA 38

935 AACCAAGTCAAGGCCATGCTATGCTAGGTCGA 970

## RESULT 15

AAS28620/c  
 ID AAS28620 standard; DNA; 14654 BP.

AC AAS28620;

07-NOV-2001 (first entry)

Genomic sequence #460 encoding for novel human respiratory antigen. Human; respiratory antigen; respiratory disorder; throat disorder; lung disorder; nose disorder; lung cancer; gene therapy; cytostatic; anti allergic; anti asthmatic; anti inflammatory; olfactory; respiratory active; ds.

Homo sapiens.

WO200155448-A1.

02-AUG-2001.

17-JAN-2001; 2001WO-US01333.

31-JAN-2000; 2000US-0179065.

04-FEB-2000; 2000US-0180828.

24-FEB-2000; 2000US-0184664.

02-MAR-2000; 2000US-0186350.

16-MAR-2000; 2000US-0189874.

17-MAR-2000; 2000US-0190076.

18-APR-2000; 2000US-0198123.

19-MAY-2000; 2000US-0205515.

07-JUN-2000; 2000US-0209467.

28-JUN-2000; 2000US-0214886.

30-JUN-2000; 2000US-0215135.

07-JUL-2000; 2000US-0215647.

PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 23-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251866.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUNA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476224/51.

Isolated polypeptide for treating, preventing and/or prognosing disorders related to the respiratory system including respiratory cancers and also for testing and detection e.g. diagnosis -

Disclosure; SED ID No 1054; 546pp; English.

The present invention relates to the isolation of novel human respiratory antigens (AAU17685-AAU17975), and cDNA and genomic sequences encoding for these polypeptides. The sequences of the invention are useful for preventing, treating and/or prognosing disorders related to the respiratory system including throat disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis), lung disorders e.g. pneumonia, allergic disorders e.g. asthma, pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of the respiratory tissues e.g. lung cancer. The polynucleotide sequences

CC of the invention are useful in gene therapy and antisense therapy.  
 CC AAS28161-AAS28764 represent genomic sequences encoding for novel  
 CC human respiratory antigens.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 14654 BP; 2492 A; 4400 C; 4798 G; 2964 T; 0 other;  
 Query Match 56.3%; Score 21.4; DB 22; Length 14654;  
 Best Local Similarity 80.6%; Pred. No. 84;  
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 8 AGGTCAAAGGTCAAAACCTAGGTCAAAGGTCA 38  
 |||||  
 DB 5366 AGGACAGAGGGCCAAAGCCAGGTCAAAGGGCA 5336

Search completed: February 19, 2003, 21:22:40  
 Job time : 29.0954 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 20:28:04 ; Search time 4.8563 Seconds  
(without alignments)  
2399.710 Million cell updates/sec

Title: US-09-808-388-2

Perfect score: 38

Sequence: 1 caaaactaggtcaaggtcaaaactaggtcaaaagggtca 38

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15339381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	57.9	5910	1	US-08-195-814-1
2	22	57.9	5910	1	US-08-195-814-1
3	20.4	53.7	3539	4	US-09-245-248B-59
4	20.4	53.7	3539	4	US-09-245-248B-53
5	20.4	53.7	4797	4	US-09-413-568F-25
6	20.4	53.7	4797	4	US-09-354-243B-25
7	20.2	53.2	11443	4	US-08-961-527-49
8	19.6	51.6	2291	3	US-08-795-430-48
9	19.6	51.6	2291	4	US-09-355-700-48
10	19.6	51.6	152331	3	US-09-128-155-16
11	19.6	51.6	176373	3	US-09-128-155-17
12	19.4	51.1	321	1	US-08-322-742-11
13	19.4	51.1	571	1	US-08-322-742-14
14	19.4	51.1	3592	3	US-08-714-948-63
15	19.4	51.1	3592	4	US-09-263-315-63
16	19.4	51.1	3592	4	US-09-263-315-63
17	19.4	51.1	3592	4	US-09-266-417-63
18	19.4	51.1	6464	1	US-08-321-478-2
19	19.4	51.1	6464	1	US-08-321-478-4
20	19.4	51.1	6464	1	US-08-321-478-6
21	19	50.0	1288	4	US-09-724-864-16
22	19	50.0	1846	4	US-09-336-536-37
23	18.8	49.5	2403	1	US-08-454-720A-41
24	18.8	49.5	3061	2	US-08-692-787-47
25	18.8	49.5	3061	4	US-09-097-199-47
26	18.8	49.5	3537	4	US-09-245-248B-58
27	18.6	48.9	122	4	US-09-437-457-3

Sequence 8, Appli  
Sequence 8, Appli  
Sequence 8, Appli  
Sequence 8, Appli  
Sequence 6, Appli  
Sequence 8, Appli  
Sequence 164, App  
Sequence 1, Appli  
Sequence 106, App  
Sequence 106, App  
Sequence 106, App  
Sequence 106, App  
Sequence 106, App  
Sequence 106, App  
Sequence 12, Appli  
Sequence 155, App  
Sequence 12, Appli  
Sequence 155, App

#### ALIGNMENTS

RESULT 1  
US-08-195-814-1  
; Sequence 1, Application US/08195814  
; Patent No. 5547869  
; GENERAL INFORMATION:  
; APPLICANT: DUMAS, BRUNO; GERVAIS, MONICA;  
; APPLICANT: BERGION, MAX; JOURDAN, MIREITTE; JOUSSET,  
; APPLICANT: FRANCOISE XAVIERE  
; TITLE OF INVENTION: NOVEL PLASMIDS  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIERMAN AND MUSERLIAN  
; STREET: 600 THIRD AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/195,814  
; FILING DATE: 14-FEB-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/881,054  
; FILING DATE: 11-MAY-1992  
; APPLICATION NUMBER: 07/278,735  
; FILING DATE: 2-DEC-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CHARLES A. MUSERLIAN  
; REGISTRATION NUMBER: 19,683  
; REFERENCE/DOCKET NUMBER: 146.1029-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 661-8000  
; TELEFAX: (212) 661-8002  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5910  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: UNKNOWN  
; TOPOLOGY: UNKNOWN  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: DENSOVIRUS  
; STRAIN: DENSOVIRUS OF JUNONIA  
; INDIVIDUAL ISOLATE:



```

1 0015
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: FLOPPY DISK
6
7 COMPUTER: IBM PC COMPATIBLE
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: WORDPERFECT 5.1
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/06/195,814
16
17 FILING DATE: 14-FEB-1994
18
19 CLASSIFICATION: 435
20
21 PRIOR APPLICATION DATA:
22
23 APPLICATION NUMBER: 07/861,054
24
25 FILING DATE: 11-MAY-1992
26
27 APPLICATION NUMBER: 07/278,735
28
29 FILING DATE: 2-DEC-1988
30
31 ATTORNEY/AGENT INFORMATION:
32
33 NAME: CHARLES A. MUSERLIAN
34
35 REGISTRATION NUMBER: 19,683
36
37 REFERENCE/POCKET NUMBER: 146,1029
38
39 TELECOMMUNICATION INFORMATION:
40
41 TELEPHONE: (212) 661-8000
42
43 TELEFAX: (212) 661-8002
44
45 INFORMATION FOR SEQ ID NO: 1:
46
47 SEQUENCE CHARACTERISTICS:
48
49 LENGTH: 5910
50
51 TYPE: NUCLEIC ACID
52

```

```

Query Match      53.7%; Score 20.4; DB 4; Length 3539;
Best Local Similarity 80.0%; Pred. No. 27;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 9 GGTCAAAGGTCAAAACCTAGGTCAAAGTCA 38
Db 3462 GGTCAAAGGTCACGCCCTACGTCTAAGTCA 3491

```

## RESULT 4

US-09-245-248B-53

; Sequence 53, Application US/09245248B

; Patent No. 6395472

; GENERAL INFORMATION:

; APPLICANT: Abbott Laboratories

; APPLICANT: Leary, Thomas

; APPLICANT: Erker, James

; APPLICANT: Chalmers, Michelle

; APPLICANT: Simons, John

; APPLICANT: Birkenmeyer, Larry

; APPLICANT: Muehrhoff, Scott

; APPLICANT: Pilot-Matias, Tami

; APPLICANT: Desai, Suresh

; APPLICANT: Mushahwar, Isha

; TITLE OF INVENTION: METHODS OF UTILIZING THE TT VIRUS

; FILE REFERENCE: 6461.US.01

; CURRENT APPLICATION NUMBER: US/09/245,248B

; CURRENT FILING DATE: 1999-02-05

; NUMBER OF SEQ ID NOS: 71

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 53

; LENGTH: 3853

; TYPE: DNA

; ORGANISM: Homo sapien

; US-09-245-248B-53

Query Match

Best Local Similarity 53.7%; Score 20.4; DB 4; Length 3853;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY

9 GGTCAAAGGTCAAAGTCAAGTCAAGGTCA 38

Db

3568 GGTCAAAGTCACTGCTAGTCAAGGTCA 3597

## RESULT 5

US-09-419-568F-25

; Sequence 25, Application US/09419568F

; Patent No. 6331613

; GENERAL INFORMATION:

; APPLICANT: Dumoutier, Laure

; APPLICANT: Louhed, Jamila

; APPLICANT: Renauld, Jean-Christophe

; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Fac

; TITLE OF INVENTION: (TIFFs) The Proteins Encoded, and Uses Thereof

; FILE REFERENCE: LUD 5543.2

; CURRENT APPLICATION NUMBER: US/09/419,568F

; CURRENT FILING DATE: 1999-10-18

; PRIOR APPLICATION NUMBER: US09/354,243

; PRIOR FILING DATE: 1999-07-16

; PRIOR APPLICATION NUMBER: US09/178,973

; PRIOR FILING DATE: 1998-10-25

; NUMBER OF SEQ ID NOS: 29

; SEQ ID NO 25

; LENGTH: 4797

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; US-09-419-568F-25

Query Match

Best Local Similarity 53.7%; Score 20.4; DB 4; Length 4797;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY

2 AAAACTAGGTCAAAGTCAAACTAGGTCA 31

Db

1788 AAATCTAGGTCACTGTTGAAATCTAGGTCA 1817

## RESULT 6

US-09-961-527-49

; Sequence 49, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 49:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11443 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-09-961-527-49

US-09-354-243B-25

; Sequence 25, Application US/09354243B

; Patent No. 6339117

; GENERAL INFORMATION:

; APPLICANT: Dumoutier, Laure

; APPLICANT: Louhed, Jamila

; APPLICANT: Renauld, Jean-Christophe

; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Fa

; TITLE OF INVENTION: (TIFFs)

; TITLE OF INVENTION: The Proteins Encoded, and Uses Thereof

; FILE REFERENCE: LUD 5543.1

; CURRENT APPLICATION NUMBER: US/09/354,243B

; CURRENT FILING DATE: 1999-07-16

; PRIOR APPLICATION NUMBER: US09/178,973

; PRIOR FILING DATE: 1998-10-25

; NUMBER OF SEQ ID NOS: 29

; SEQ ID NO 25

; LENGTH: 4797

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; US-09-354-243B-25

Query Match

Best Local Similarity 53.7%; Score 20.4; DB 4; Length 4797;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY

2 AAAACTAGGTCAAAGTCAAACTAGGTCA 31

Db

1788 AAATCTAGGTCACTGTTGAAATCTAGGTCA 1817

## RESULT 7

US-08-961-527-49

; Sequence 49, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 49:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11443 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-08-961-527-49

Query Match 53.2%; Score 20.2; DB 4; Length 11443;  
 Best Local Similarity 75.8%; Pred. No. 38;  
 Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 4 AACTAGGTCAAGGTCAAACTAGGTCAAGGT 36  
 Db 6901 AAAAGGTCAAGTACCAAACTGGATTAAAGGT 6933

RESULT 8  
 US-08-795-430-48/c  
 ; Sequence 48, Application US/08795430  
 ; Patent No. 6130071  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alitalo, Karl  
 ; JOUKOV, Vladimir  
 ; TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)  
 ; TITLE OF INVENTION: Protein and Gene, Mutants Thereof, and Uses Thereof  
 ; NUMBER OF SEQUENCES: 57  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/795,430  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/FI96/00427  
 ; FILING DATE: 01-AUG-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/671,573  
 ; FILING DATE: 28-JUN-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/601,132  
 ; FILING DATE: 14-FEB-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/585,895  
 ; FILING DATE: 12-JAN-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/340,011  
 ; FILING DATE: 14-NOV-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gass, David A.  
 ; REGISTRATION NUMBER: 38,153  
 ; REFERENCE/DOCKET NUMBER: 28967/33691  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 48:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2991 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-795-430-48

Query Match 51.6%; Score 19.6; DB 3; Length 2991;  
 Best Local Similarity 73.5%; Pred. No. 52;  
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 CAAAGTCTAGGTCAAAAGTCAAACTAGGTCAAG 34  
 Db 1968 CAAAGTTTGGAAAAGGTCACTACTATGTGACAG 1935  
 RESULT 9  
 US-09-355-700-48/c  
 ; Sequence 48, Application US/09355700  
 ; Patent No. 6361946  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ludwig Institute for Cancer Research  
 ; Helsinki University Licensing  
 ; Alitalo, Karl (U.S. only)  
 ; JOUKOV, Vladimir (U.S. only)  
 ; TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)  
 ; TITLE OF INVENTION: Protein and Gene, Mutants Thereof, and Uses Thereof  
 ; NUMBER OF SEQUENCES: 59  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/355,700  
 ; FILING DATE: 05-NOV-1999  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/795,430  
 ; FILING DATE: 05-FEB-1997  
 ; APPLICATION NUMBER: PCT/FI96/00427  
 ; FILING DATE: 01-AUG-1996  
 ; APPLICATION NUMBER: 08/671,573  
 ; FILING DATE: 28-JUN-1996  
 ; APPLICATION NUMBER: 08/601,132  
 ; FILING DATE: 14-FEB-1996  
 ; APPLICATION NUMBER: 08/585,895  
 ; FILING DATE: 12-JAN-1996  
 ; APPLICATION NUMBER: 08/510,133  
 ; FILING DATE: 01-AUG-1995  
 ; APPLICATION NUMBER: 08/340,011  
 ; FILING DATE: 14-NOV-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gass, David A.  
 ; REGISTRATION NUMBER: 38,153  
 ; REFERENCE/DOCKET NUMBER: 28967/34140  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 48:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2991 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 48:  
 ; US-09-355-700-48

Query Match 51.6%; Score 19.6; DB 4; Length 2991;  
 Best Local Similarity 73.5%; Pred. No. 52;  
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 CAAAGTCTAGGTCAAAAGTCAAACTAGGTCAAG 34

Db 1968 CAAAGTTTGGAAAAAGGTCATCATGTGTGACAG 1935

RESULT 10  
US-09-128-155-16/c  
; Sequence 16, Application US/09128155  
; Patent No. 6117654  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/09/128,155  
; EARLIER FILING DATE: 1998-08-03  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER FILING DATE: 1997-08-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 16  
; LENGTH: 152331  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(152331)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-16

Query Match 51.6%; Score 19.6; DB 3; Length 152331;  
Best Local Similarity 73.5%; Pred. No. 94;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 5 ACTAGTCAAGGTCAAAGCTAGGTCAAAGGTCA 38  
Db 143694 ACTAGTGTAGCCAGAGCTAGAGCCAGGTCA 143661

RESULT 11  
US-09-128-155-17  
; Sequence 17, Application US/09128155  
; Patent No. 6117654  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/09/128,155  
; EARLIER FILING DATE: 1998-08-03  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER FILING DATE: 1997-08-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 176373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(176373)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-17

Query Match 51.6%; Score 19.6; DB 3; Length 176373;  
Best Local Similarity 73.5%; Pred. No. 96;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 5 ACTAGTCAAGGTCAAAGCTAGGTCAAAGGTCA 38

Db 50297 ACTAGTGTATAGCCAGAGCTAGAGCCAGGTCA 50330

RESULT 12  
US-08-322-742-11  
; Sequence 11, Application US/08322742  
; Patent No. 5688641  
; GENERAL INFORMATION:  
; APPLICANT: Sager, Ruth  
; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
; COMPUTER: IBM PS/2 Model 50Z or 55SX  
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/322,742  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/938,823  
; FILING DATE: September 1, 1992  
; APPLICATION NUMBER: 07/844,296  
; FILING DATE: February 28, 1992  
; APPLICATION NUMBER: 07/552,216  
; FILING DATE: February 28, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frazer, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 00530/048003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 321  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-322-742-11

Query Match 51.1%; Score 19.4; DB 1; Length 321;  
Best Local Similarity 70.3%; Pred. No. 43;  
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 1 CAAAACCTAGGTCAAAGGTCAAAGCTAGGTCAAAGGTC 37  
Db 129 CAAGTAAAGTCAAGGCCAAGCCAGTCAGTCAGGTC 165

RESULT 13  
US-08-322-742-14  
; Sequence 14, Application US/08322742  
; Patent No. 5688641  
; GENERAL INFORMATION:  
; APPLICANT: Sager, Ruth  
; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 50Z or 55SX  
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/322,742  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/938,823  
FILING DATE: September 1, 1992  
APPLICATION NUMBER: 07/844,296  
FILING DATE: February 28, 1992  
APPLICATION NUMBER: 07/552,216  
FILING DATE: February 28, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frazer, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 00530/048003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 571  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-322-742-14

Query Match 51.1%; Score 19.4; DB 1; Length 571;  
Best Local Similarity 70.3%; Pred. No. 47;  
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAGGTCAAACTAGGTCAAGGTC 37  
DB 184 CAAGATAAAGTCAAGGCCAAGCCAGTCAAGGTC 220

RESULT 14  
US-08-714-918-63  
Sequence 63, Application US/08714918  
Patent No. 6037123  
GENERAL INFORMATION:  
APPLICANT: Benton, Bret  
APPLICANT: Lee, Ving J.  
APPLICANT: Malouin, Francois  
APPLICANT: Martin, Patrick K.  
APPLICANT: Schmid, Molly B.  
APPLICANT: Sun, Dongxu  
TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL  
TITLE OF INVENTION: TARGET GENES  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/714,918  
FILING DATE: September 13, 1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,102  
FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 222/005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3592 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-714-918-63

Query Match 51.1%; Score 19.4; DB 3; Length 3592;  
Best Local Similarity 69.7%; Pred. No. 63;  
Matches 23; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 6 CTAGGTCAAGGTCAAACTAGGTCAAGGTC 38  
DB 1591 CTAGGTAAATGTCMGATTTAGATCAAAATCTTA 1623

RESULT 15  
US-09-265-315-63  
Sequence 63, Application US/09265315  
Patent No. 6187541  
GENERAL INFORMATION:  
APPLICANT: Benton, Bret  
APPLICANT: Lee, Ving J.  
APPLICANT: Malouin, Francois  
APPLICANT: Martin, Patrick K.  
APPLICANT: Schmid, Molly B.  
APPLICANT: Sun, Dongxu  
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
TITLE OF INVENTION: TARGET GENES  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/265,315  
FILING DATE: March 9, 1999  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/714,918  
FILING DATE: September 13, 1996  
APPLICATION NUMBER: 60/009,102  
FILING DATE: December 22, 1995  
APPLICATION NUMBER: 60/003,798  
FILING DATE: September 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 240/247

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3592 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-265-315-63

```

```

Query Match      51.1%; Score 19.4; DB 4; Length 3592;
Best Local Similarity 69.7%; Pred. No. 63;
Matches 23; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

```

```

2Y      6 CTAGGTCAAAGGTCAAACTAGGTCAAAGGTCA 38
      ||||| ||||| ||||| ||||| ||||| |||||
Db 1591 CTAGGTAAATGTCMGAWTTAGTCAATCTTA 1623

```

```

Search completed: February 20, 2003, 01:13:38
Job time : 55.8563 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

M nucleic - nucleic search, using sw model

run on: February 19, 2003, 23:17:14 ; Search time 28.8245 Seconds  
(without alignments)  
671.453 Million cell updates/sec

title: US-09-808-388-2  
perfect score: 38  
sequence: 1 caaaactaggtcaagggtcaaaactagggtcaagggtca 38

scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

searched: 424239 seqs, 254661826 residues

total number of hits satisfying chosen parameters: 848478

minimum DB seq length: 0

maximum DB seq length: 200000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

result No.	Score	Query Match	Length DB	ID	Description
1	38	100.0	38	10	US-09-808-388-2
2	26	68.4	60	9	Sequence 2, Appli
3	26	68.4	60	9	Sequence 142, App
4	25	65.8	41	10	US-09-877-738A-142
5	25	65.8	332	10	Sequence 3, Appli
6	21.6	56.8	397658	10	Sequence 6, Appli
7	21.4	56.3	14654	10	Sequence 3, Appli
8	21	55.3	75899	10	Sequence 1054, App
9	20.4	53.7	3539	10	Sequence 243, App
10	20.4	53.7	3853	10	Sequence 59, Appl
11	20.4	53.7	4797	10	Sequence 53, Appl
12	20.2	53.2	52	10	Sequence 25, Appl
13	20.2	53.2	94	10	Sequence 4, Appli
14	20.2	53.2	480	10	Sequence 355, App
15	20.2	53.2	3697	12	Sequence 10963, A
16	20.2	53.2	55155	10	Sequence 58, Appl
17	20	52.6	20	9	Sequence 3, Appli
18	20	52.6	20	9	Sequence 67, Appl
19	20	52.6	20	9	Sequence 68, Appl
					Sequence 67, Appl

US-09-877-738A-68  
Sequence 1, Appli  
Sequence 1941, App  
Sequence 292, App  
Sequence 529, App  
Sequence 3950, App  
Sequence 4951, App  
Sequence 5734, App  
Sequence 2937, App  
Sequence 3, Appli  
Sequence 2, Appli  
Sequence 16, Appli  
Sequence 17, Appli  
Sequence 1104, App  
Sequence 211, App  
Sequence 483, App  
Sequence 3918, App  
Sequence 1126, App  
Sequence 1798, App  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 4248, App  
Sequence 15751, A  
Sequence 3237, App  
Sequence 1042, App  
Sequence 582, App

US-09-808-388-2  
Sequence 2, Application US/09808388  
Patent No. US20020081719A1  
GENERAL INFORMATION:  
APPLICANT: Massaad, Charbel  
APPLICANT: Berenbaum, Francis  
APPLICANT: Olivier, Jean-Luc  
APPLICANT: Salvat, Colette  
APPLICANT: Berezat, Gilbert  
TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
FILE REFERENCE: ST00010  
CURRENT APPLICATION NUMBER: US/09/808,388  
CURRENT FILING DATE: 2001-09-20  
PRIOR APPLICATION NUMBER: FR/00/03262  
PRIOR FILING DATE: 2000-03-14  
PRIOR APPLICATION NUMBER: US 60/196,959  
PRIOR FILING DATE: 2000-04-13  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 38  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: PPR element  
US-09-808-388-2

Query Match 100.0%; Score 38; DB 10; Length 38;  
Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAAGGTCAAAACTAGGTCAAGGTCA 38  
Db 1 CAAACTAGGTCAAGGTCAAAACTAGGTCAAGGTCA 38

RESULT 2  
US-09-877-705A-142/c  
Sequence 142, Application US/09877705A  
Publication No. US20030008283A1

## ALIGNMENTS

RESULT 1  
US-09-808-388-2  
Sequence 2, Application US/09808388  
Patent No. US20020081719A1  
GENERAL INFORMATION:  
APPLICANT: Massaad, Charbel  
APPLICANT: Berenbaum, Francis  
APPLICANT: Olivier, Jean-Luc  
APPLICANT: Salvat, Colette  
APPLICANT: Berezat, Gilbert  
TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
FILE REFERENCE: ST00010  
CURRENT APPLICATION NUMBER: US/09/808,388  
CURRENT FILING DATE: 2001-09-20  
PRIOR APPLICATION NUMBER: FR/00/03262  
PRIOR FILING DATE: 2000-03-14  
PRIOR APPLICATION NUMBER: US 60/196,959  
PRIOR FILING DATE: 2000-04-13  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 38  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: PPR element  
US-09-808-388-2

Query Match 100.0%; Score 38; DB 10; Length 38;  
Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAAGGTCAAAACTAGGTCAAGGTCA 38  
Db 1 CAAACTAGGTCAAGGTCAAAACTAGGTCAAGGTCA 38

RESULT 2  
US-09-877-705A-142/c  
Sequence 142, Application US/09877705A  
Publication No. US20030008283A1





```
SEQ ID NO 3
LENGTH: 397658
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(397658)
OTHER INFORMATION: n = A,T,C or G
IS-09-813-320-3

Query Match          56.8%; Score 21.6; DB 10; Length 397658;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Y 2 AAAAAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 37
b 292121 AAGACTCGTAAAGTTCAAGACTAGGAGGCAAGTC 292155

RESULT 7
US-09-764-860-1054/c
Sequence 1054, Application US/09764860
Patent No. US20020094953A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P0008
CURRENT APPLICATION NUMBER: US/09/764,860
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1198
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1054
LENGTH: 14654
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-860-1054

Query Match          56.3%; Score 21.4; DB 10; Length 14654;
Best Local Similarity 80.6%; Pred. No. 38;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Y 8 AGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 38
b 5366 AGGACAGGCGCAAGCCAGGTCAGGTCAGGTC 5336

RESULT 8
US-09-854-883-243/c
Sequence 243, Application US/09854883
Patent No. US20020055479A1
GENERAL INFORMATION:
APPLICANT: Lex M. Cowser
APPLICANT: Jacqueline Wyatt
APPLICANT: Susan M. Freier
APPLICANT: Brett P. Monia
APPLICANT: Madeline M. Butler
APPLICANT: Robert McKay
TITLE OF INVENTION: ANTISENSE MODULATION OF PTPIB EXPRESSION
FILE REFERENCE: ISPH-0576
CURRENT APPLICATION NUMBER: US/09/854,883
CURRENT FILING DATE: 2001-05-14
PRIOR APPLICATION NUMBER: US 09/629,644
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 09/487,368
PRIOR FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 389
SEQ ID NO 243
LENGTH: 75899
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-854-883-243

Query Match          55.3%; Score 21; DB 10; Length 75899;
Best Local Similarity 73.0%; Pred. No. 71;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Y 2 AAAAAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 38
b 17097 AACAGTAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 17061

RESULT 9
US-09-815-656-59
Sequence 59, Application US/09815656
Patent No. US20010041331A1
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Leary, Thomas
APPLICANT: Erker, James
APPLICANT: Chalmers, Michelle
APPLICANT: Simons, John
APPLICANT: Birkenmeyer, Larry
APPLICANT: Muerhoff, Scott
APPLICANT: Pilot-Matias, Tami
APPLICANT: Desai, Suresh
APPLICANT: Mushahwar, Isa
TITLE OF INVENTION: METHODS OF UTILIZING THE TT VIRUS
FILE REFERENCE: 6461.US.01
CURRENT APPLICATION NUMBER: US/09/815,656
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: 09/245,248
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 71
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 59
LENGTH: 3539
TYPE: DNA
ORGANISM: Homo sapien
US-09-815-656-59

Query Match          53.7%; Score 20.4; DB 10; Length 3539;
Best Local Similarity 80.0%; Pred. No. 71;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Y 9 GGTCAAAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 38
b 3462 GGTCAAAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 3491

RESULT 10
US-09-815-656-53
Sequence 53, Application US/09815656
Patent No. US20010041331A1
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Leary, Thomas
APPLICANT: Erker, James
APPLICANT: Chalmers, Michelle
APPLICANT: Simons, John
APPLICANT: Birkenmeyer, Larry
APPLICANT: Muerhoff, Scott
APPLICANT: Pilot-Matias, Tami
APPLICANT: Desai, Suresh
APPLICANT: Mushahwar, Isa
TITLE OF INVENTION: METHODS OF UTILIZING THE TT VIRUS
FILE REFERENCE: 6461.US.01
CURRENT APPLICATION NUMBER: US/09/815,656
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: 09/245,248
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 71
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 53
LENGTH: 3853
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-815-656-53

Query Match      53.7%; Score 20.4; DB 10; Length 3853;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 9 GGTCAGAGGTCAAAAGTCTAGGTCAAGGTCA 38
   |||||
Db 3568 GGTCAGAGGTCAAGTCTAGGTCAAGGTCA 3597

RESULT 11
US-09-751-797-25
; Sequence 25, Application US/09751797
; Patent No. US20010024652A1
; GENERAL INFORMATION:
; APPLICANT: Dumoutier, Laure
; APPLICANT: Louhed, Jamila
; APPLICANT: Renauld, Jean-Christophe
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Fac
; TITLE OF INVENTION: (Tifs) The Proteins Encoded, and Uses Thereof
; FILE REFERENCE: LUD 5543.2
; CURRENT APPLICATION NUMBER: US/09/751,797
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: 09/419,568
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: US09/178,973
; PRIOR FILING DATE: 1998-10-26
; NUMBER OF SEQ ID NOS: 29
; SEQ ID NO 25
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-751-797-25

Query Match      53.7%; Score 20.4; DB 10; Length 4797;
Best Local Similarity 80.0%; Pred. No. 74;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AAAACTAGGTCAAAAGTCAAAAGTCAAGTCA 31
   |||||
Db 1788 AAACTAGGTCACTGTGAAATCTAGGTCA 1817

RESULT 12
US-09-808-388-4
; Sequence 4, Application US/09808389
; Patent No. US20020081719A1
; GENERAL INFORMATION:
; APPLICANT: Massaad, Charbel
; APPLICANT: Berenbaum, Francis
; APPLICANT: Olivier, Jean-Luc
; APPLICANT: Salvat, Colette
; APPLICANT: Berezat, Gilbert
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them
; TITLE OF INVENTION: their uses
; FILE REFERENCE: ST00010
; CURRENT APPLICATION NUMBER: US/09/808,388
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: FR/00/03262
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: US 60/196,959
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 52
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: PPRE element
US-09-808-388-4

Query Match      53.2%; Score 20.2; DB 10; Length 52;
Best Local Similarity 88.0%; Pred. No. 41;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 14 AAGGTCAAAAGTCTAGGTCAAAAGTCA 38
   |||||
Db 28 AGCCCAAAAGTCTAGGTCAAAAGTCA 52

RESULT 13
US-09-294-093B-395/c
; Sequence 395, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalquidi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 395
; LENGTH: 94
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700342675H1
; NAME/KEY: unsure
; LOCATION: 62, 70-71, 84
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-395

Query Match      53.2%; Score 20.2; DB 10; Length 94;
Best Local Similarity 75.8%; Pred. No. 45;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CAAAAGTCTAGGTCAAAAGTCAAAAGTCTCAAA 33
   |||||
Db 34 CAATCAAGGTAAATTTCAATCAAGGTAAAA 2

RESULT 14
US-09-864-761-10963/c
; Sequence 10963, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
```

PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 10963  
LENGTH: 480  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC005277.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.92  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.79  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.63  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.79  
JS-09-864-761-10963

Query Match 53.2%; Score 20.2; DB 10; Length 480;  
Best Local Similarity 75.8%; Pred. No. 60;  
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

2y 2 AAAACTAGGTCAAAGGTCAAAAGTCAAGTCAAG 34  
db 239 ATAATGAGGTCCTAGGAGAACTTAGGTCAAAG 207

RESULT 15  
US-10-002-600-58  
Sequence 58, Application US/10002600  
Patent No. US20020137077A1  
GENERAL INFORMATION:  
APPLICANT: Hopkins, Christopher M.  
APPLICANT: Peterson, David P.  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Hawkins, Phillip R.  
TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS  
FILE REFERENCE: PA-0042 US  
CURRENT APPLICATION NUMBER: US/10/002,600  
CURRENT FILING DATE: 2001-10-25  
PRIOR APPLICATION NUMBER: 60/243,521  
PRIOR FILING DATE: 2000-10-25  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PERL Program  
SEQ ID NO 58  
LENGTH: 3697  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:

NAME/KEY: misc feature  
OTHER INFORMATION: Template ID: 039976.16  
NAME/KEY: unsure  
LOCATION: 1444, 1446, 1448-1449, 1454, 1456  
OTHER INFORMATION: a, t, c, g, or other  
US-10-002-600-58  
Query Match 53.2%; Score 20.2; DB 12; Length 3697;  
Best Local Similarity 75.8%; Pred. No. 84;  
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
2y 2 AAAACTAGGTCAAAGGTCAAAAGTCAAGTCAAG 34  
db 852 AAATCTAGATCAAAAGAGCCTAATCTAAGCCCAAG 884  
Search completed: February 20, 2003, 06:47:25  
Job time : 127.825 secs

GenCore version 5.1.3  
Copyright (C) 1993 - 2003 Compugen Ltd.

WM nucleic - nucleic search, using sw model  
run on: February 19, 2003, 19:33:48 ; Search time 184.898 Seconds  
(without alignments)  
3328.484 Million cell updates/sec

Title: us-09-808-388-2  
Perfect score: 38  
Sequence: 1 caaaactaggtcaagggtcaaaactagggtcaaaagtca 38

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:
- 1: em\_estba:\*\*
  - 2: em\_esthum:\*\*
  - 3: em\_estin:\*\*
  - 4: em\_estmu:\*\*
  - 5: em\_estov:\*\*
  - 6: em\_estpl:\*\*
  - 7: em\_estro:\*\*
  - 8: em\_hcc:\*\*
  - 9: gb\_est1:\*\*
  - 10: gb\_est2:\*\*
  - 11: gb\_hcc:\*\*
  - 12: gb\_est3:\*\*
  - 13: gb\_est4:\*\*
  - 14: gb\_est5:\*\*
  - 15: em\_estfun:\*\*
  - 16: em\_estom:\*\*
  - 17: gb\_gss:\*\*
  - 18: em\_gss\_hum:\*\*
  - 19: em\_gss\_inv:\*\*
  - 20: em\_gss\_pln:\*\*
  - 21: em\_gss\_vrt:\*\*
  - 22: em\_gss\_fun:\*\*
  - 23: em\_gss\_mam:\*\*
  - 24: em\_gss\_mus:\*\*
  - 25: em\_gss\_other:\*\*
  - 26: em\_gss\_pro:\*\*
  - 27: em\_gss\_rtd:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26.4	69.5	301	9 AV165535	AV165535 AV165535
C 2	26.4	69.5	307	10 AW457734	AW457734 UI-M-BH3-
C 3	26.4	69.5	429	10 BB690786	BB690786 BB690786
C 4	26.4	69.5	453	12 BF470083	BF470083 UI-M-BH3-
C 5	26.4	69.5	454	10 AW493749	AW493749 UI-M-BH3-
C 6	26.4	69.5	454	10 AW494112	AW494112 UI-M-BH3-

C 7	26.4	69.5	469	14 BM894724	BM894724 ih70g02.y
C 8	26.4	69.5	477	9 AA144171	AA144171 mg54g11.x
C 9	26.4	69.5	500	10 BB754909	BB754909 BB754909
C 10	26.4	69.5	503	10 BE491339	BE491339 uc52h10.y
C 11	26.4	69.5	517	9 AA276552	AA276552 uc42g10.y
C 12	26.4	69.5	557	10 AW492996	AW492996 UI-M-BH3-
C 13	26.4	69.5	573	9 AA956334	AA956334 UI-R-E0-B
C 14	26.4	69.5	575	12 BG347085	BG347085 gac87f10.
C 15	26.4	69.5	606	13 BM238433	BM238433 K0519A09-
C 16	26.4	69.5	626	10 BE372290	BE372290 601223416
C 17	26.4	69.5	680	13 BM239965	BM239965 K0543B05-
C 18	26.4	69.5	741	10 BE283191	BE283191 601103773
C 19	26.4	69.5	749	12 BG671888	BG671888 DRNEZB11
C 20	26.4	69.5	795	13 B1153324	B1153324 602917187
C 21	26.4	69.5	954	9 AU079043	AU079043 AU079043
C 22	24.2	63.7	501	12 BF807899	BF807899 RC3-C1004
C 23	23.8	62.6	146	12 BF807907	BF807907 RC3-C1004
C 24	23.6	62.1	277	14 T89672	T89672 yd99d06.s1
C 25	23.6	62.1	283	9 AV291197	AV291197 AV291197
C 26	23.6	62.1	343	13 BM149493	BM149493 TCAP2Q97
C 27	23.6	62.1	351	10 AW802788	AW802788 IL2-UM007
C 28	23.6	62.1	369	12 BF401889	BF401889 UI-R-CAL-
C 29	23.6	62.1	388	9 AI844119	AI844119 UI-M-ALL-
C 30	23.6	62.1	398	10 BB836089	BB836089 BB836089
C 31	23.6	62.1	401	9 AI425917	AI425917 wg20g09.x
C 32	23.6	62.1	401	9 AL121170	AL121170 DKF2D7620
C 33	23.6	62.1	410	12 BF404793	BF404793 UI-R-CAL-
C 34	23.6	62.1	415	10 BE244244	BE244244 TCAP2PE09
C 35	23.6	62.1	423	10 BE244214	BE244214 TCAP2PE09
C 36	23.6	62.1	429	10 AW238234	AW238234 xp23c05.x
C 37	23.6	62.1	436	14 W88983	W88983 zh70f01.s1
C 38	23.6	62.1	442	10 BE334532	BE334532 us80e01.y
C 39	23.6	62.1	456	13 B1848816	B1848816 471355 MA
C 40	23.6	62.1	460	10 BE095053	BE095053 CM3-BT079
C 41	23.6	62.1	463	9 AI604273	AI604273 mtlc02.y
C 42	23.6	62.1	464	9 AA847803	AA847803 od30g01.s
C 43	23.6	62.1	469	10 AW824668	AW824668 us14e07.x
C 44	23.6	62.1	471	12 BB819560	BB819560 RC3-BN033
C 45	23.6	62.1	472	9 AI360345	AI360345 gy89g01.x

ALIGNMENTS

RESULT 1  
AV165535/c  
LOCUS  
DEFINITION  
AV165535 Mus musculus head C57BL/6J 13-day embryo Mus musculus CDNA  
clone 3110037124, mRNA sequence.  
ACCESSION  
AV165535  
VERSION  
AV165535.1 GI:5371972  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 301)  
AUTHORS  
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, P., Iahikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Mitsuhashi, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.





Query Match 69.5%; Score 26.4; DB 12; Length 453;  
 Best Local Similarity 96.4%; Pred. No. 38;  
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 TCAAGGTCAAAGTCTAGGTCAAAGGTCA 38  
 |||||  
 DB 368 TCAAGGTCAAAGTCTAGGTCAAAGGTCA 341  
 |||||

RESULT 5  
 AW493749  
 LOCUS  
 DEFINITION UI-M-BH3-aui-g-04-0-UI.s1 NIH\_BMAP\_M.S4 Mus musculus cDNA clone  
 UI-M-BH3-aui-g-04-0-UI 3', mRNA sequence.

ACCESSION AW493749.1 GI:7064030  
 VERSION  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 454)  
 Bonaldo,M.F.; Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 JOURNAL 97044477  
 MEDLINE  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mestr@mail.nih.gov  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 normalized olfactory bulbs library cDNA Library Preparation: M.B.  
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENETICS. It should be noted that Bento Soares  
 is generating a small number of additional specialized  
 non-redundant arrays of BMAP cDNAs whose availability will be  
 considered under appropriate and limited collaborative arrangements  
 Seq primer: M13 Forward  
 POLYA=Yes.

FEATURES Location/Qualifiers  
 1..454  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-aui-g-04-0-UI"  
 /clone\_lib="NIH\_BMAP\_M.S4"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; The  
 NIH\_BMAP\_M.S4 library is a subtracted library of a series,  
 ultimately derived from a mixture of individually tagged  
 normalized libraries from ten regions of the mouse brain  
 (cerebellum, brain stems, olfactory bulbs, hypothalamus,  
 cortex, amygdala, basal ganglia, pineal gland, striatum,  
 hippocampus) after a series of subtractions to reduce the  
 representation of cDNAs from which ESTs had already been  
 generated. The following serially subtracted libraries  
 were generated in this process: NIH\_BMAP\_M.S4,  
 NIH\_BMAP\_M.S3.3, NIH\_BMAP\_M.S3.2, NIH\_BMAP\_M.S3.1,  
 NIH\_BMAP\_M.S2, NIH\_BMAP\_M.S1. The subtracted library  
 (NIH\_BMAP\_M.S4) was constructed as follows: PCR amplified  
 cDNA inserts from NIH\_BMAP\_M.S3.3, NIH\_BMAP\_M.S3.2 and  
 NIH\_BMAP\_M.S3.1 clones from which 3' ESTs had been derived

was used as a driver in a hybridization with a pool of  
 the NIH\_BMAP\_M.S3.3, NIH\_BMAP\_M.S3.2, and NIH\_BMAP\_M.S3.1  
 libraries in the form of single-stranded circles. The  
 remaining single-stranded circles (subtracted library)  
 was purified by hydroxyapatite column chromatography,  
 converted to double-stranded circles and electroporated  
 into DH10B bacteria (Life Technologies) to generate the  
 NIH\_BMAP\_M.S4 library. This procedure has been previously  
 described (Bonaldo, Lennon and Soares, Genome Research  
 6:791-806, 1996)  
 TAG\_LIB=NIH\_BMAP\_M.S4  
 TAG\_TISSUE=olfactory-bulbs  
 TAG\_SEQ=CATGG  
 BASE COUNT 153 a 96 g 130 t  
 ORIGIN  
 Query Match 69.5%; Score 26.4; DB 10; Length 454;  
 Best Local Similarity 96.4%; Pred. No. 38;  
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 TCAAGGTCAAAGTCTAGGTCAAAGGTCA 38  
 |||||  
 DB 278 TCAAGGTCAAAGTCTAGGTCAAAGGTCA 305  
 |||||

RESULT 6  
 AW494112  
 LOCUS  
 DEFINITION UI-M-BH3-aui-g-09-0-UI.s1 NIH\_BMAP\_M.S4 Mus musculus cDNA clone  
 UI-M-BH3-aui-g-09-0-UI 3', mRNA sequence.

ACCESSION AW494112  
 VERSION AW494112.1 GI:7064393  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 454)  
 Bonaldo,M.F.; Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 JOURNAL 97044477  
 MEDLINE  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mestr@mail.nih.gov  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 normalized olfactory bulbs library cDNA Library Preparation: M.B.  
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENETICS. It should be noted that Bento Soares  
 is generating a small number of additional specialized  
 non-redundant arrays of BMAP cDNAs whose availability will be  
 considered under appropriate and limited collaborative arrangements  
 Seq primer: M13 Forward  
 POLYA=Yes.

FEATURES Location/Qualifiers  
 1..454  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-aui-g-09-0-UI"  
 /clone\_lib="NIH\_BMAP\_M.S4"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1; Not 1; Site 2; Eco RI; The NIH\_BMAP\_M\_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged (normalized) libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH\_BMAP\_M\_S4, NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1, NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library (NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH\_BMAP\_M\_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:1791-806, 1996).

NIH LIB=NIH\_BMAP\_M\_S4  
TAG\_TISSUE=Olfactory-bulbs  
TAG\_SEQ=CATGG

ASE COUNT 152 a 76 c 97 g 129 t  
RIGIN

Query Match 69.5%; Score 26.4; DB 10; Length 454;  
Best Local Similarity 96.4%; Pred. No. 38;  
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 11 TCAAAGTCAAACTAGGTCAAAGGTCA 38  
|||||  
b 278 TCAAAGTCAAACTAGGTCAAAGGTCA 305

RESULT 7  
M894724/c  
OCUS  
DEFINITION  
B894724  
CCESION  
ERSION  
EYWORDS  
OURCE  
ORGANISM  
REFERENCE  
AUTHORS

M894724 469 bp mRNA linear EST 28-MAR-2002  
cDNA clone IMAGE:5681858 5', mRNA sequence.  
B894724  
B894724.1 GI:19350192  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 469)  
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Blistain, A.,  
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas  
M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T.,  
Jackson, Y. and Bowers, Y.  
M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T.,  
Jackson, Y. and Bowers, Y.  
Unpublished (2000)

Endocrine Pancreas Consortium  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biochem.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:  
Washington University Genome Sequencing Center This clone is  
available royalty-free through LLNL; please contact the IMAGE  
consortium (info@image.llnl.gov) for further information

Seq primer: -40RP from Gibco  
High quality sequence stop: 432.  
Location/Qualifiers

source

1. .469  
/organism="Mus musculus"  
/strains="ICR"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5681858"  
/clone\_lib="Melton Mouse E16 5 Pancreas Library 2 M16B2"  
/sex="Both"  
/tissue\_type="Total pancreas"  
/dev\_stage="Embryonic day 16.5"  
/lab\_host="TOP10"

/note="Organ: Pancreas; Vector: pBluescript II SK; Site 1:  
Not1; Site 2: SalI; Library constructed using SuperScript  
Plasmid Library kit (Life Technologies). cDNA made by  
oligo-dT priming. Size-selected by column fractionation;  
average insert size 1.08kb. Primary library,  
unamplified."

BASE COUNT 115 a 110 c 81 g 163 t  
ORIGIN

Query Match 69.5%; Score 26.4; DB 14; Length 469;  
Best Local Similarity 96.4%; Pred. No. 38;  
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 TCAAAGTCAAACTAGGTCAAAGGTCA 38  
|||||  
Db 260 TCAAAGTCAAACTAGGTCAAAGGTCA 233

RESULT 8  
AA144171/c  
LOCUS

DEFINITION  
mq54g11.r1 Soares thymus\_2NbMT Mus musculus cDNA clone IMAGE:582596  
5', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA144171  
AA144171.1 GI:1713539  
EST.  
house mouse.  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 477)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

The WashU-HMI Mouse EST Project  
Unpublished (1996)

CONTACT: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:357244

Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 420.

Location/Qualifiers

1. .477  
/organism="Mus musculus"  
/strains="CS7BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:582596"  
/clone\_lib="Soares\_thymus\_2NbMT"  
/sex="male"  
/tissue\_type="Thymus"  
/dev\_stage="4 weeks"



/lab\_host="DH10B"  
 /note=vector: pTV73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5,  
 TGTACCAATCTGAAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pTV73 vector. RNA  
 provided by Dr. Bertrand Jordan. Library went through two  
 rounds of normalization, and was constructed by Bento  
 Soares and M.Fatima Bonaldo."  
 Soares and M.Fatima Bonaldo."

BASE COUNT 129 a 108 c 84 g 156 t

ORIGIN

Query Match 69.5%; Score 26.4; DB 9; Length 477;  
 Best Local Similarity 96.4%; Pred. No. 39;  
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 TCAAGGTCAAACTAGTCAAAAGGTCA 38

Db 246 TCAAGGTCAAACTAGTCAAAAGGTCA 219

RESULT 9  
 BB754909/c  
 LOCUS  
 DEFINITION BB754909 RIKEN full-length enriched, melanocyte Mus musculus cDNA  
 clone G270045K23 3', mRNA sequence.

ACCESSION BB754909  
 VERSION BB754909  
 KEYWORDS EST.  
 SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 500)  
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
 Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii  
 Y., Ito,M., Kawai,J., Kojima,Y., Konno.H., Kouda,M., Matsuyama,T.,  
 Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T.,  
 Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
 Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,  
 A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
 Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
 2001)

JOURNAL

COMMENT  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp/  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
 N., Konno.H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 Wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
 Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
 Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10  
 (11), 1757-1771 (2000)  
 Konno.H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
 Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.  
 e mouse tissues.

FEATURES

source

1..500  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="G270045K23"  
 /clone\_lib="RIKEN full-length enriched, melanocyte"  
 /cell\_type="melanocyte"  
 /note="pooled tissues; (tissue type=cerebellum,  
 dev\_stage=16 days neonate, sex=mixed).  
 (tissue type=cerebellum, dev\_stage=0 day neonate,  
 sex=mixed), (tissue type=hippocampus, dev\_stage=adult,  
 sex=male), (tissue type=whole body, dev\_stage=9 days  
 embryo, sex=mixed), (tissue type=lung, dev\_stage=13 days  
 embryo, sex=mixed)"

BASE COUNT 134 a 115 c 84 g 167 t

ORIGIN

Query Match 69.5%; Score 26.4; DB 10; Length 500;  
 Best Local Similarity 96.4%; Pred. No. 39;  
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 TCAAGGTCAAACTAGTCAAAAGGTCA 38

Db 226 TCAAGGTCAAACTAGTCAAAAGGTCA 199

RESULT 10  
 BE449139/c

LOCUS

DEFINITION BE449139 503 bp mRNA linear EST 25-JUL-2000  
 ut52h10.y1 Soares mouse 3nbms Mus musculus cDNA clone IMAGE:3331555  
 5', mRNA sequence.

ACCESSION BE449139

VERSION BE449139.1 GI:9448716

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 503)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

TITLE

JOURNAL

COMMENT

Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:1075719

Seq primer: -40RP from Gibco

High quality sequence stop: 467.

Location/Qualifiers

FEATURES

source

1..503  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3331555"  
 /clone\_lib="Soares mouse 3nbms"  
 /sex="male"  
 /tissue\_type="Spleen"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"

/note=vector: pTV73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5,  
 TGTACCAATCTGAAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pTV73 vector. RNA  
 provided by Dr. Bertrand Jordan. Library went through

Matches 27; Conservative 0; Mismatches 1

was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH-BMAP M54 library. This procedure has been previously described (Bonaldo, Lemmon and Soares, Genome Research 6:791-806, 1996)  
 TAG LIB=NIH-BMAP M54  
 TAG TISSUE=glfactory-bulbs  
 TAG SEQ=CATGG

BASE COUNT 184 a 123 g 153 t  
 ORIGIN

Query Match 69.5%; Score 26.4; DB 10; Length 557;  
 Best Local Similarity 96.4%; Pred. No. 41;  
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 TCAAAGGTCAAAGTCTAGGTCAAAGGTCA 38  
 Db 278 TCAAAGGTCAAAGTCTAGGTCAAAGGTCA 305

RESULT 13  
 LOCUS

AA859634 573 bp mRNA linear EST 03-JUL-1999  
 DEFINITION UI-R-E0-bs-h-10-0-UI-s1 UI-R-E0 Rattus norvegicus cDNA clone  
 UI-R-E0-bs-h-10-0-UI-3' similar to gb|U69546|HSU69546 Human RNA  
 binding protein Etr-3 mRNA, complete cds, mRNA sequence.

ACCESSION AA859634  
 VERSION AA859634.1 GI:4230179  
 KEYWORDS EST.

SOURCE

ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 573)

AUTHORS Bonaldo,M.F., Lemmon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 9704477

COMMENT On Mar 10, 1998 this sequence version replaced gi:2949154.

CONTACT Soares, MB

PROGRAM for Rat Gene Discovery and Mapping

UNIVERSITY of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

TEL: 319 335 8250

FAX: 319 335 9565

EMAIL: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 18-Day-Embryo library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1777095

Seq primer: M13 Forward

POLYA-No.

FEATURES

SOURCE

Location/Qualifiers

1..573  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-E0-bs-h-10-0-UI"  
 /clone\_lib="UI-R-E0"  
 /dev\_stage="embryonic"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: NotI; Site\_2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows

identification of the library of origin of a clone within the mixture."

BASE COUNT 121 a 148 c 98 g 206 t  
 ORIGIN

Query Match 69.5%; Score 26.4; DB 9; Length 573;

Best Local Similarity 96.4%; Pred. No. 42;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 TCAAAGGTCAAAGTCTAGGTCAAAGGTCA 38

Db 502 TCAAAGGTCAAAGTCTAGGTCAAAGGTCA 475

RESULT 14

LOCUS

EG347085 575 bp mRNA linear EST 28-FEB-2001  
 DEFINITION dac87f10.V1 Wellcome CRC pcDNA1 St24-26 Xenopus laevis cDNA clone  
 IMAGE:4437258 5', mRNA sequence.

ACCESSION EG347085

VERSION EG347085.1 GI:13167509

KEYWORDS EST.

SOURCE

ORGANISM African clawed frog.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 575)

AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,

Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,

B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,

Waterston,R. and Wilson,R.

WashU Xenopus EST project, 1999

UNPUBLISHED (1999)

CONTACT: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

TEL: 314 286 1800

FAX: 314 286 1810

EMAIL: estwatson.wustl.edu

Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.

Gurdon (Wellcome/CRC Institute). DNA sequencing by: Washington

University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov

Seq primer: -40RP from Gibco

High quality sequence stop: 472.

FEATURES

SOURCE

Location/Qualifiers

1..575  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:4437258"  
 /clone\_lib="Wellcome CRC pcDNA1 St24-26"  
 /tissue\_type="pooled embryos, stage 24-26"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Vector: pcDNA1; Site\_1: NotI; Site\_2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Newkooop and Faber. Library was constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute)."

BASE COUNT 151 a 126 c 112 g 186 t

ORIGIN

Query Match 69.5%; Score 26.4; DB 12; Length 575;

Best Local Similarity 96.4%; Pred. No. 42;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 TCAAAGGTCAAAGTCTAGGTCAAAGGTCA 38

Db 334 TCAAAGGTCAAAGTCTAGGTCAAAGGTCA 307

```

RESULT 15
BM238433      606 bp      mRNA      linear      EST 31-JAN-2002
LOCUS      K0519A09-3 NTA Mouse Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1+)
DEFINITION      cDNA Library (Long) Mus musculus cDNA clone K0519A09 3', mRNA
sequence.
ACCESSION      BM238433
VERSION
KEYWORDS
SOURCE      EST.
SOURCE      BM238433.1 GI:17873799
ORGANISM      house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 606)
Pao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G.,
Taub, D., Longo, D.L., Keller, J. and Ko, M.S.H.
Systematic Analyses of NIA Mouse Hematopoietic Stem Cell
(Lin-/c-Kit+/Sca-1+) cDNA Library (Long)
JOURNAL      Unpublished (2001)
COMMENT      Laboratory of Genetics
Contact: Dawood B. Dudekula
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: K0519 row: A column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 606
POLYA=Yes.

FEATURES
    source
        1..606
            /organism="Mus musculus"
            /strain="C57BL/6Ncr"
            /db_xref="niaEST:K0519A09-3"
            /db_xref="taxon:10090"
            /clone="K0519A09"
            /clone_lib="NIA Mouse Hematopoietic Stem Cell
            (Lin-/c-Kit+/Sca-1+) cDNA Library (Long)"
            /tissue_type="Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1+
            )"
            /dev_stage="Age approx. 10 weeks old"
            /lab_host="DH10B"
            /note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
            NotI; Mouse cDNA project by the Laboratory of Genetics,
            National Institute on Aging (NIA), Intramural Research
            Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is
            a long-transcript enriched cDNA library (Ref. Genome Res.
            11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
            obtained from Drs. Dennis Taub, Dan Longo (National
            Institute on Aging, USA), Jonathan Keller (National Cancer
            Institute, USA). Double-stranded cDNAs were synthesized
            with an Oligo(dT) primer [Invitrogen:
            5'-pGACTAGTCTAGACGCGAGCGCCGCCCTTTT-3'] from
            4.8 ug of total RNA, treated with T4 DNA polymerase, and
            purified by ethanol-precipitation. The cDNAs were ligated
            to Lona-linker Lb-Sal4, purified by phenol/chloroform, and
            separated from free linkers by Centricon 100. Then, the
            cDNAs were amplified by long-range high fidelity PCR using
            Ex Taq polymerase (Takara) with a primer Sal4-S. The
            products were purified by phenol/chloroform and Centricon
            100. The cDNAs were digested with SalI and NotI enzymes
            and cloned into SalI/NotI site of pSPORT1 plasmid vector.
            The DH10B E. coli host was transformed with the ligation
            mixture by the standard chemical method. The average
            insert size is about 2.7 kb. The library was constructed
            by Yulan Piao (NIA)."
BASE COUNT      195 a      113 c      134 g      164 t
ORIGIN
Query Match      69.5%; Score 26.4; DB 13; Length 606;
Best Local Similarity 96.4%; Pred. No. 43;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2Y      11 TCAAAGGTCAAAACCTAGGTCAAGGTCA 38

```

```

|||||
Db      275 TCAAAGGTCAAAACCTAGGTCAAGGTCA 302
|||||

```

Search completed: February 20, 2003, 01:08:48  
Job time : 186.898 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

WM nucleic - nucleic search, using sw model  
Run on: February 19, 2003, 19:28:33 ; Search time 127.515 Seconds  
(without alignments)  
9357.426 Million cell updates/sec  
Title: US-09-808-388-3  
Perfect score: 41  
Sequence: 1 caaaactaggtcaaaagtca.....caaaactaggtcaaaagtca 41  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_man.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	41	100.0	41	6	AX251575	AX251575 Sequence
2	41	100.0	332	6	AX251578	AX251578 Sequence
3	26.8	65.4	69	6	AX063386	AX063386 Sequence
4	26.2	63.9	199866	3	AY003872	AY003872 Plasmodiu
5	25	61.0	38	6	AX251574	AX251574 Sequence
C 6	24.8	60.5	27567	3	AF040658	AF040658 Caenorhab
7	24.8	60.5	154455	2	AC021844	AC021844 Homo sapi
C 8	24.8	60.5	155394	2	AC092328	AC092328 Homo sapi
9	24.8	60.5	184635	9	AC025445	AC025445 Homo sapi
C 10	24.8	60.5	250178	2	AC006771	AC006771 Caenorhab
C 11	24.8	60.5	299727	2	AC006738	AC006738 Caenorhab
12	24.6	60.0	50530	9	AL365187	AL365187 Human DNA
C 13	24.6	60.0	172454	2	AC025008	AC025008 Homo sapi
C 14	24.2	59.0	72012	2	AC023759	AC023759 Homo sapi
C 15	24	58.5	14056	1	SPDEXCAP	247210 S.pneumonia
C 16	24	58.5	35842	3	CEZK228	282086 Caenorhabdi
C 17	24	58.5	135693	9	HSJ1167H4	AL121914 Human DNA
C 18	24	58.5	184533	9	AC093592	AC093592 Homo sapi
C 19	24	58.5	192573	9	AC073344	AC073344 Homo sapi
C 20	23.8	58.0	141025	2	AC123528	AC123528 Oryza sat
21	23.8	58.0	169298	2	AC098240	AC098240 Rattus no
C 22	23.8	58.0	202565	9	AL354696	AL354696 Human DNA
C 23	23.6	57.6	15450	14	SV41ENA	X64275 Simian viru
C 24	23.6	57.6	18306	9	AC125334	AC125334 Homo sapi
C 25	23.6	57.6	110000	2	CEY39B6.2	Continuation (3 of
C 26	23.6	57.6	143255	2	AC015996	AC015996 Homo sapi
C 27	23.6	57.6	177556	2	AC021433	AC021433 Homo sapi
28	23.6	57.6	246237	3	CEY39B6A	AL132948 Caenorhab
29	23.2	56.6	78135	2	AC096187	AC096187 Rattus no
C 30	23.2	56.6	92458	9	AL353715	AL353715 Human DNA
C 31	23.2	56.6	173791	9	AC021035	AC021035 Homo sapi
C 32	23.2	56.6	182442	2	AC128099	AC128099 Rattus no
C 33	23	56.1	8160	1	U39681	U39681 Mycoplasma
C 34	23	56.1	65489	2	AC102506	AC102506 Mus muscu
C 35	23	56.1	145709	1	D90914	D90914 Synchocyst
36	23	56.1	160071	2	AL844607	AL844607 Mus muscu
37	23	56.1	176780	2	AC130427	AC130427 Homo sapi
38	23	56.1	178093	2	AC016530	AC016530 Homo sapi
C 39	23	56.1	187251	2	AC069233	AC069233 Homo sapi
C 40	23	56.1	188682	2	AC124825	AC124825 Mus muscu
C 41	22.8	55.6	38505	3	CEZK930	Z70213 Caenorhabdi
42	22.8	55.6	196496	2	AL805910	AL805910 Mus muscu
C 43	22.8	55.6	225537	2	AC099613	AC099613 Mus muscu
C 44	22.8	55.6	249285	2	AC093482	AC093482 Mus muscu
C 45	22.6	55.1	39726	3	CEB0413	Z92824 Caenorhabdi

## ALIGNMENTS

RESULT 1  
AX251575  
LOCUS AX251575 41 bp DNA linear PAT 05-OCT-2001  
DEFINITION Sequence 3 from Patent WO0168845.  
ACCESSION AX251575  
VERSION AX251575.1 GI:15984998  
KEYWORDS .  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 41)  
AUTHORS Masaaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.  
TITLE Inflammation-inducible hybrid promoters, vectors containing same  
and uses thereof  
JOURNAL Patent: WO 0168845-A 3 20-SEP-2001;

```

Aventis Pharma S.A. (FR)
Location/Qualifiers
1. .41
/organism="synthetic construct"
/db xref="taxon:32630"
/nc="element PPRE"
18 a 8 c 8 g 7 t

Query Match 100.0%; Score 41; DB 6; Length 41;
Best Local Similarity 100.0%; Pred. No. 8.7e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAAACCTAGGTCAGGTCATCAAACTAGGTCAGGTCAGGTC 41
Db 1 CAAAACCTAGGTCAGGTCATCAAACTAGGTCAGGTCAGGTC 41

RESULT 2
AX251578 332 bp DNA linear PAT 05-OCT-2001
LOCUS
DEFINITION
Sequence 6 from Patent WO0168845.
ACCESSION
AX251578
VERSION
AX251578.1 GI:15985001
KEYWORDS
synthetic construct.
synthetic construct.
artificial sequences.
REFERENCE
1 (bases 1 to 332)
Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.
Inflammation-inducible hybrid promoters, vectors containing same
and uses thereof
JOURNAL
Patent: WO 0168845-A 6 20-SEP-2001;
Aventis Pharma S.A. (FR)
Location/Qualifiers
1. .332
/organism="synthetic construct"
/db xref="taxon:32630"
/nc="promoteur hybride PPRE/PLA2s"
96 a 91 c 82 g 63 t

Query Match 100.0%; Score 41; DB 6; Length 332;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAAACCTAGGTCAGGTCATCAAACTAGGTCAGGTCAGGTC 41
Db 13 CAAAACCTAGGTCAGGTCATCAAACTAGGTCAGGTCAGGTC 53

RESULT 3
AX063386 69 bp DNA linear PAT 24-JAN-2001
LOCUS
DEFINITION
Sequence 6 from Patent WO0078986.
ACCESSION
AX063386
VERSION
AX063386.1 GI:12541176
KEYWORDS
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 69)
Darteil,R., Crouzet,J., Staels,B. and Mahfoudi,A.
Regulation system of expression using nuclear ppar receptors
Patent: WO 0078986-A 6 28-DEC-2000;
Aventis Pharma S.A. (FR)
Location/Qualifiers
1. .69
/organism="Homo sapiens"
/db xref="taxon:9606"
27 a 14 c 17 g 11 t

Query Match 65.4%; Score 26.8; DB 6; Length 69;
Best Local Similarity 92.9%; Pred. No. 15;
Matches 39; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 CAAAACCTAGGTCAGGTCATCAAACTAGGTCAGGTCAGGTC 41
Db 17 CAAAACCTAGGTCAGGTCACGGAACTAGGTCAGGTCAGGTC 58

RESULT 4
AY003872 199866 bp DNA linear INV 05-DEC-2001
LOCUS
Plasmodium vivax YAC 1H14, complete sequence.
DEFINITION
AY003872
ACCESSION
AY003872
VERSION
AY003872.1 GI:14578280
KEYWORDS
Plasmodium vivax.
Plasmodium vivax.
ORGANISM
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 (bases 1 to 199866)
Tchavtchitch,M., Fischer,K., Huestis,R. and Saul,A.
The sequence of a 200 Kb portion of a Plasmodium vivax chromosome
reveals a high degree of conservation with Plasmodium falciparum
chromosome 3
JOURNAL
Mol. Biochem. Parasitol. 118 (2), 211-222 (2001)
11738711
REFERENCE
2 (bases 1 to 199866)
Tchavtchitch,M., Fischer,K., Huestis,R. and Saul,A.
Direct Submission
Submitted (21-JUN-2000) Malaria Biology Laboratory, The Queensland
Institute of Medical Research, PO Royal Brisbane Hospital,
Brisbane, Queensland Q4029, Australia
JOURNAL
3 (bases 1 to 199866)
Tchavtchitch,M., Fischer,K., Huestis,R. and Saul,A.
Direct Submission
Submitted (24-AUG-2001) Malaria Biology Laboratory, The Queensland
Institute of Medical Research, PO Royal Brisbane Hospital,
Brisbane, Queensland Q4029, Australia
JOURNAL
REMARK
Amino acid sequence updated by submitter
FEATURES
Location/Qualifiers
1..199866
/organism="plasmodium vivax"
/db xref="taxon:5855"
/chromosome="7"
/clone="YAC 1H14"
/country="Brazil; Porto Velho"
/note="isolated from erythrocytes"
<1..>1540
/gene="PVLH14005w"
/join(<186..923,1058..>1540)
/gene="PVLH14005w"
/product="PVLH1405_P"
/join(<186..923,1058..>1540)
/gene="PVLH14005w"
/note="hypothetical protein"
/codon_start=1
/product="PVLH1405_P"
/protein_id="AAF99447.2"
/db xref="GI:15284207"
/translation="DVFQDLVDYQIADYIKTFNFGYGTIKCDKFNSDDEVEPSIK
RICNVDFRLKATVLTAVNTVVKTKYHVELYNLWLANRELRILKDRTRFREPKYIE
TNSMVDLEGALKNTYDLNVEVENMILHLKYDYVKIISKEKNKNSKYSEEND
LYKGMKKCYETRFNFEYNALKNLSYKNVYNSNLCKEILPOLPILATNEKSK
KFMQKTVESCDMLKNKDNSEFPQRYKYNILGLTAQEQYKILSTNAELPLCSKYC
GSIIPFDEQNGKLKTLCAKFANNLINLSDNLQNVESACSCSYFTYTYIKIMNM
FNNKTSVYFPHSILNGLNEVLYQVNSLPLVGAKSCLFYLDGNFEKVEEKLYLHDFKN
YDQISKQGNCKY"
<10125..>11171
/gene="PVLH14010w"
/join(<10125..10179,10314..>10897,10998..>11171)
/gene="PVLH14010w"
/product="PVLH14010_P"

```



Db 198931 CAAAAAAGGCGCAAGGTATTCAAAAAAGGCGCAAGGT 198969

RESULT 5  
 AX251574 LOCUS AX251574 38 bp DNA linear PAT 05-OCT-2001  
 DEFINITION Sequence 2 from Patent WO0168845.  
 ACCESSION AX251574  
 VERSION AX251574.1 GI:15984997  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM  
 REFERENCE 1 (bases 1 to 38)  
 AUTHORS Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Berezat,G.  
 TITLE Inflammation-inducible hybrid promoters, vectors containing same and uses thereof  
 JOURNAL Patent: WO 0168845-A 2 20-SEP-2001;  
 Aventis Pharma S.A. (FR)  
 FEATURES  
 source 1..38  
 /organism="synthetic construct"  
 /db xref="taxon:32630"  
 /note="element PPR2"  
 BASE COUNT 17 a 7 c 8 g 6 t  
 ORIGIN  
 Query Match 61.0%; Score 25; DB 6; Length 38;  
 Best Local Similarity 92.7%; Pred. No. 73;  
 Matches 38; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
 QY 1 CAAACTAGGTCRAAGTCATCAAACTAGTCAGTCAAGGTCA 41  
 Db 1 CAAACTAGGTCRAAGG---TCAAACTAGTCAGTCAAGGTCA 38  
 RESULT 6  
 AF040658/C LOCUS AF040658 27567 bp DNA linear INV 24-MAY-2002  
 DEFINITION Caenorhabditis elegans cosmid W07G9, complete sequence.  
 ACCESSION AF040658  
 VERSION AF040658.2 GI:14625307  
 KEYWORDS HTG.  
 SOURCE  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Peloderae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 27567)  
 AUTHORS Waterston,R.  
 TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium  
 JOURNAL Science 282 (5396), 2012-2018 (1998)  
 MEDLINE 99069613  
 PUBMED 9851916  
 REFERENCE 2 (bases 1 to 27567)  
 AUTHORS Pauley,A., Goela,D., Le,T.T. and Wilson,R.  
 TITLE The sequence of C. elegans cosmid W07G9  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 27567)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-DEC-1997) Department of Genetics, Washington University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 4 (bases 1 to 27567)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JUL-2001) Department of Genetics, Washington University, Genome Sequencing Center, 444 Forest Park Avenue, St. Louis, MO 63110, USA  
 REFERENCE 5 (bases 1 to 27567)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-2002) Department of Genetics, Washington

COMMENT

University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
 On Jul 6, 2001 this sequence version replaced gi:2746905.

Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RQ, England  
 email: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=W07G9;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is Y37E11AR, 600 bp overlap; the 3' cosmid is K09B3, 1000 bp overlap. Actual start of this cosmid is at base position 1 of W07G9; actual end is at 5949 of M02B7.

NOTES:

Coding sequences below are the result of integration and manual review of the following data : computer analysis using the program Gensfinder (P. Green and L. Hillier, personal communication), the large scale EGR projects of Yuji Kohara ([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C. elegans ORFeome cloning project (<http://wormfdb.fci.harvard.edu/>), similarity to other proteins from BlastX analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source 1..27567  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db xref="taxon:6239"  
 /chromosome="IV"  
 /clone="W07G9"  
 complement(18588..19095)  
 /gene="W07G9.1"  
 /note="For a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=W07G9.1;class=Sequence>"  
 complement(join(18588..18663,18733..18848,18930..18995,19054..19095))  
 /gene="W07G9.1"  
 /codon\_start=1  
 /product="Hypothetical protein W07G9.1"  
 /protein\_id="AAB95056.1"  
 /db xref="GI:2746906"  
 /translation="MGGLVSNFYGIFVSSNSSTTLVDCKYNAIVEQTILEFV EFSNSIIFLNLYSKVMNNSNFKKFFRKNSVTYVSNFLFSKNLYKFGV"  
 BASE COUNT 9030 a 4374 c 4800 g 9363 t

Query Match 60.5%; Score 24.8; DB 3; Length 27567;



Best Local Similarity 80.6%; Pred. No. 53;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

2y 2 AAAAAGTGGTCAAGGTCATCAAAAGTGGTCAAG 37  
|||||  
Db 23614 AAAAAGTGGTCAAGGTCATCAAAAGTGGTCAAG 23579  
|||||

RESULT 7  
AC021844  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone RP11-15D23, WORKING DRAFT SEQUENCE,  
4 unordered pieces.  
ACCESSION AC021844  
VERSION AC021844.4 GI:7534251  
KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Waterston, R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 154455)  
Waterston, R.H.  
Direct Submission  
Submitted (20-JAN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Apr 11, 2000 this sequence version replaced gi:7523912.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H NH0015D23  
----- Summary Statistics -----  
Sequencing vector: M13, 100%  
Sequencing method: plasmid; 0%  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 152020 bases at least Q40  
Consensus quality: 152947 bases at least Q30  
Consensus quality: 153535 bases at least Q20  
Insert size: 164000; agarose-fp  
Insert coverage: 5.45 in Q20 bases; agarose-fp  
Quality coverage: 5.80 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 10635: contig of 10635 bp in length  
\* 10636 10735: gap of unknown length  
\* 10736 44171: contig of 33436 bp in length  
\* 44172 44271: gap of unknown length  
\* 44272 98542: contig of 54271 bp in length  
\* 98543 98642: gap of unknown length  
\* 98643 154455: contig of 55813 bp in length.  
Location/Qualifiers  
1..154455  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-15D23"

misc\_feature 1..10635  
/note="assembly\_name:Contig3  
clone\_end:17  
vector\_side:right"  
10736..44171  
/note="assembly\_name:Contig4"  
44272..98542  
/note="assembly\_name:Contigs"  
98643..154455  
/note="assembly\_name:Contig6  
clone\_end:SP6  
vector\_side:left"  
BASE COUNT 48764 a 27957 c 28022 g 49405 t 307 others  
ORIGIN

Query Match 60.5%; Score 24.8; DB 2; Length 154455;  
Best Local Similarity 80.6%; Pred. No. 47;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CAAAAGTGGTCAAGGTCATCAAAAGTGGTCAAA 36  
Db 146934 CAAAAGTGGTCAAGGTCATCAAAAGTGGTCAAA 146969  
|||||

RESULT 8  
AC092328/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone RP11-15D23, WORKING DRAFT SEQUENCE,  
2 unordered pieces.  
ACCESSION AC092328  
VERSION AC092328.2 GI:15290443  
KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 155394)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Aug 25, 2001 this sequence version replaced gi:14589517.

----- Genome Center -----  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 412220  
Center clone name: RPCI-11\_15D23  
-----  
Summary Statistics  
Consensus quality: 153679 bases at least Q40  
Consensus quality: 153819 bases at least Q30  
Consensus quality: 153879 bases at least Q20  
Estimated insert size: 169360; agarose-fp estimation  
Estimated insert size: 155294; sum-of-contigs estimation  
Quality coverage: 12.32 in Q20 bases; agarose-fp estimation  
Quality coverage: 13.43 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 25817: contig of 25817 bp in length  
\* 25818 25917: gap of unknown length

FEATURES  
source

\* 25918 155394: contig of 129477 bp in length.

FEATURES  
source

Location/Qualifiers  
1..155394  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-15D23"  
/clone.lib="RP11-15D23"  
/clone.lib="RP11-15D23"  
BASE COUNT 46482 a 28012 c 28437 g 52363 t 100 others  
ORIGIN

Query Match 60.5%; Score 24.8; DB 2; Length 155394;  
Best Local Similarity 80.8%; Pred. No. 47;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCACAAAGTCATCAAACTAGGTCAAA 36

Db 107089 CAAATTAAGCAAAAGTCACATACAGAGTCAAA 107054

RESULT 9  
AC025445

LOCUS Homo sapiens chromosome 5 clone CTD-2049017, complete sequence.  
DEFINITION Homo sapiens chromosome 5 clone CTD-2049017, complete sequence.  
ACCESSION AC025445  
VERSION AC025445.5 GI:15451698  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 184635)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 184635)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS Direct Submission  
TITLE Unpublished  
JOURNAL  
REFERENCE 2 (bases 1 to 184635)  
DOE Joint Genome Institute.  
AUTHORS Direct Submission  
TITLE Submitted (09-MAR-2000) Production Sequencing Facility, DOE Joint  
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 184635)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS Direct Submission  
TITLE Submitted (04-OCT-2000) DOE Joint Genome Institute, 2800 Mitchell  
JOURNAL Drive, Walnut Creek, CA 94598, USA  
REFERENCE 4 (bases 1 to 184635)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS Direct Submission  
TITLE Submitted (06-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell  
JOURNAL Drive, Walnut Creek, CA 94598, USA

COMMENT On Sep 6, 2001 this sequence version replaced gi:10567849.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.4% of Sequence;  
Estimated Total Number of Errors is 0.8.  
STS Content:  
SHGC-15011 G17012.

FEATURES  
source

Location/Qualifiers  
1..184635  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2049017"  
BASE COUNT 62006 a 33523 c 33180 g 55926 t  
ORIGIN

Query Match 60.5%; Score 24.8; DB 9; Length 184635;  
Best Local Similarity 80.6%; Pred. No. 46;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCACAAAGTCATCAAACTAGGTCAAA 36

Db 28636 CAAATAAGACAAAGTCACATACAGAGTCAAA 28671

RESULT 10  
AC006771/c

LOCUS Caenorhabditis elegans clone Y46C8, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
DEFINITION Caenorhabditis elegans clone Y46C8, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
ACCESSION AC006771  
VERSION AC006771.1 GI:4263167  
KEYWORDS HTG, HTGS, PHASE1  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 250178)  
Waterston,R.H.  
TITLE The sequence of Caenorhabditis elegans clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 250178)  
Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

COMMENT \* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 3214: contig of 3214 bp in length  
\* 3215  
\* 3227: gap of unknown length  
\* 3228  
\* 7785: contig of 4558 bp in length  
\* 7786  
\* 7799: gap of unknown length  
\* 13501  
\* 13503: contig of 5702 bp in length  
\* 13504  
\* 20142: contig of 6629 bp in length  
\* 20155  
\* 20156: gap of unknown length  
\* 20157  
\* 27438: contig of 7283 bp in length  
\* 27439  
\* 27452: gap of unknown length  
\* 27453  
\* 37159: gap of unknown length  
\* 37171  
\* 46384: contig of 9813 bp in length  
\* 46385  
\* 46397: gap of unknown length  
\* 46398  
\* 250178: contig of 203181 bp in length.

FEATURES  
source

Location/Qualifiers  
1..250178  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/clone="Y46C8"  
BASE COUNT 78496 a 45960 c 46480 g 79151 t 91 others  
ORIGIN

Query Match 60.5%; Score 24.8; DB 2; Length 250178;  
Best Local Similarity 80.6%; Pred. No. 45;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AAAACTAGGTCACAAAGTCATCAAACTAGGTCAAG 37

Db 107223 AAAACTAGGTCACAAAGTCATCAAACTAGGTCAAG 107188

RESULT 11  
AC006738/c

LOCUS Caenorhabditis elegans clone Y37E11a, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION Caenorhabditis elegans clone Y37E11a, \*\*\* SEQUENCING IN PROGRESS  
ACCESSION AC006738  
VERSION AC006738.1 GI:4263200

Query Match 60.5%; Score 24.8; DB 2; Length 250178;  
Best Local Similarity 80.6%; Pred. No. 45;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AAAACTAGGTCACAAAGTCATCAAACTAGGTCAAG 37

Db 107223 AAAACTAGGTCACAAAGTCATCAAACTAGGTCAAG 107188

RESULT 11  
AC006738/c

LOCUS Caenorhabditis elegans clone Y37E11a, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION Caenorhabditis elegans clone Y37E11a, \*\*\* SEQUENCING IN PROGRESS  
ACCESSION AC006738  
VERSION AC006738.1 GI:4263200

Query Match 60.5%; Score 24.8; DB 9; Length 184635;  
Best Local Similarity 80.6%; Pred. No. 46;  
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCACAAAGTCATCAAACTAGGTCAAA 36



Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 3, 2000 this sequence version replaced gi:7145052.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

COMMENT

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WTBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
Center project name: L7735  
Center clone name: 223 G 4

----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 164665 bases at least Q40  
Consensus quality: 168473 bases at least Q30  
Consensus quality: 169984 bases at least Q20

Insert size: 170000; agarose-fp  
Insert size: 171054; sum-of-contigs  
Quality coverage: 4.9 in Q20 bases; agarose-fp  
Quality coverage: 4.9 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. \* as soon as it is available and the accession number will be preserved.

1 574: contig of 574 bp in length  
\* 575 674: gap of 100 bp  
\* 675 3496: contig of 2822 bp in length  
\* 3497 3596: gap of 100 bp  
\* 3597 6375: contig of 2779 bp in length  
\* 6376 6475: gap of 100 bp  
\* 6476 9360: contig of 2885 bp in length  
\* 9361 9460: gap of 100 bp  
\* 9461 12588: contig of 3128 bp in length  
\* 12589 12688: gap of 100 bp  
\* 12689 16911: contig of 4223 bp in length  
\* 16912 17011: gap of 100 bp  
\* 17012 22194: contig of 5183 bp in length  
\* 22195 22294: gap of 100 bp  
\* 22295 31113: contig of 8819 bp in length  
\* 31114 31213: gap of 100 bp  
\* 31214 40652: contig of 9439 bp in length  
\* 40653 40752: gap of 100 bp  
\* 40753 52339: contig of 11587 bp in length  
\* 52340 52439: gap of 100 bp  
\* 52440 74858: contig of 22419 bp in length  
\* 74859 74958: gap of 100 bp  
\* 74959 95195: contig of 20237 bp in length  
\* 95196 95295: gap of 100 bp  
\* 95296 117436: contig of 22141 bp in length  
\* 117437 117536: gap of 100 bp  
\* 117537 142063: contig of 24527 bp in length  
\* 142064 142163: gap of 100 bp  
\* 142164 172454: contig of 30291 bp in length.

FEATURES  
source

Location/Qualifiers  
1..172454  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1"  
/clone="RP11-223G4"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..574  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
675..3496  
/note="assembly\_fragment"  
3597..6375  
/note="assembly\_fragment"  
6476..9360  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:left  
9461..12588  
/note="assembly\_fragment"  
12689..16911  
/note="assembly\_fragment"  
17012..22194  
/note="assembly\_fragment"  
22295..31113  
/note="assembly\_fragment"  
31214..40652  
/note="assembly\_fragment"  
40753..52339  
/note="assembly\_fragment"  
52440..74858  
/note="assembly\_fragment"  
74959..95195  
/note="assembly\_fragment"  
95296..117436  
/note="assembly\_fragment"  
117537..142063  
/note="assembly\_fragment"  
142164..172454  
/note="assembly\_fragment"

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

Query Match 60.0%; Score 24.6; DB 2; Length 172454;  
Best Local Similarity 76.9%; Pred. No. 55;  
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 3 AAACATCTCAAGGTCATCAAACTAGGTCAAGGTCA 41

Db 16025 AAACATCTCAAGGTCATCAAACTAGGTCAAGGTCA 15987

RESULT 14

AC023759/c 72012 bp DNA linear HTG 13-JUL-2000  
LOCUS  
DEFINITION Homo sapiens clone RP11-11N3, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC023759  
VERSION AC023759.2 GI:7144826  
KEYWORDS HTG; HTGS\_PHASE0.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 72012)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE

Homo sapiens, clone RP11-11N3

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 72012)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,  
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,

Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
DeArelano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,  
Fenster, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hages, B., Heaford, A., Horton, L.,  
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Landers, T., Laroque, K., Lenockzy, J., Levine, R.,  
Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,  
McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M.,  
Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,  
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,  
Severy, P., Spencer, J.B., Stange-Romann, N., Stojanovic, N.,  
Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,  
Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,  
Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and  
Zody, M.

# TITLE JOURNAL

## COMMENT

Direct Submission  
Submitted (17-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 3, 2000 this sequence version replaced gi:6984406.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L3235  
Center clone name: 11\_N\_3

-----  
\* NOTE: This record contains 81 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be Gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 753: contig of 753 bp in length  
754 853: gap of 100 bp  
854 1664: contig of 811 bp in length  
1665 1764: gap of 100 bp  
1765 2582: contig of 818 bp in length  
2583 2682: gap of 100 bp  
2683 3466: contig of 784 bp in length  
3467 3566: gap of 100 bp  
3567 4350: contig of 784 bp in length  
4351 4450: gap of 100 bp  
4451 5223: contig of 773 bp in length  
5224 5323: gap of 100 bp  
5324 6097: contig of 774 bp in length  
6098 6197: gap of 100 bp  
6198 7002: contig of 805 bp in length  
7003 7102: gap of 100 bp  
7103 7901: contig of 799 bp in length  
7902 8001: gap of 100 bp  
8002 8786: contig of 785 bp in length  
8787 8886: gap of 100 bp  
8887 9691: contig of 805 bp in length  
9692 9791: gap of 100 bp  
9792 10576: contig of 785 bp in length  
10577 10676: gap of 100 bp  
10677 11460: contig of 784 bp in length  
11461 11560: gap of 100 bp  
11561 12344: contig of 784 bp in length  
12345 12444: gap of 100 bp

12445 13239: contig of 795 bp in length  
13240 13339: gap of 100 bp  
13340 14100: contig of 761 bp in length  
14101 14200: gap of 100 bp  
14201 14990: contig of 790 bp in length  
14991 15090: gap of 100 bp  
15091 15878: contig of 788 bp in length  
15879 15978: gap of 100 bp  
15979 16763: contig of 785 bp in length  
16764 16863: gap of 100 bp  
16864 17663: contig of 800 bp in length  
17665 17763: gap of 100 bp  
17764 18545: contig of 782 bp in length  
18546 18645: gap of 100 bp  
18646 19447: contig of 802 bp in length  
19448 19477: gap of 100 bp  
19478 20336: contig of 789 bp in length  
20337 20436: gap of 100 bp  
20437 21253: contig of 817 bp in length  
21254 21353: gap of 100 bp  
21354 22112: contig of 759 bp in length  
22113 22212: gap of 100 bp  
22213 22969: contig of 757 bp in length  
22970 23069: gap of 100 bp  
23070 23853: contig of 784 bp in length  
23854 23953: gap of 100 bp  
23954 24741: contig of 788 bp in length  
24742 24841: gap of 100 bp  
24842 25644: contig of 803 bp in length  
25645 25744: gap of 100 bp  
25745 26526: contig of 782 bp in length  
26527 26626: gap of 100 bp  
26627 27417: contig of 791 bp in length  
27418 27517: gap of 100 bp  
27518 28332: contig of 815 bp in length  
28333 28432: gap of 100 bp  
28434 29246: contig of 814 bp in length  
29247 29346: gap of 100 bp  
29347 30134: contig of 788 bp in length  
30135 30234: gap of 100 bp  
30235 31004: contig of 770 bp in length  
31005 31104: gap of 100 bp  
31105 31885: contig of 781 bp in length  
31886 31985: gap of 100 bp  
31986 32762: contig of 777 bp in length  
32763 32862: gap of 100 bp  
32863 33672: contig of 810 bp in length  
33673 33772: gap of 100 bp  
33773 34542: contig of 770 bp in length  
34543 34642: gap of 100 bp  
34643 35421: contig of 779 bp in length  
35422 35521: gap of 100 bp  
35522 36336: contig of 815 bp in length  
36337 36436: gap of 100 bp  
36437 37210: contig of 774 bp in length  
37211 37310: gap of 100 bp  
37311 38120: contig of 810 bp in length  
38121 38220: gap of 100 bp  
38221 39033: contig of 813 bp in length  
39034 39133: gap of 100 bp  
39134 39942: contig of 809 bp in length  
39943 40042: gap of 100 bp  
40043 40823: contig of 787 bp in length  
40824 40929: gap of 100 bp  
40930 41720: contig of 791 bp in length  
41721 41820: gap of 100 bp  
41821 42618: contig of 798 bp in length  
42619 42718: gap of 100 bp  
42719 43486: contig of 768 bp in length  
43487 43586: gap of 100 bp  
43587 44335: contig of 809 bp in length  
44336 44435: gap of 100 bp  
44436 45276: contig of 781 bp in length

\* 45277 45376: gap of 100 bp  
 \* 45377 46162: contig of 786 bp in length  
 \* 46163 46262: gap of 100 bp  
 \* 46263 47034: contig of 772 bp in length  
 \* 47035 47134: gap of 100 bp  
 \* 47135 47902: contig of 768 bp in length  
 \* 47903 48002: gap of 100 bp  
 \* 48003 48797: contig of 795 bp in length  
 \* 48798 48897: gap of 100 bp  
 \* 48898 49699: contig of 802 bp in length  
 \* 49700 49799: gap of 100 bp  
 \* 49800 50579: contig of 780 bp in length  
 \* 50580 50679: gap of 100 bp  
 \* 50680 51465: contig of 786 bp in length  
 \* 51466 51565: gap of 100 bp  
 \* 51566 52370: contig of 805 bp in length  
 \* 52371 52470: gap of 100 bp  
 \* 52471 53270: contig of 800 bp in length  
 \* 53271 53370: gap of 100 bp  
 \* 53371 54165: contig of 795 bp in length  
 \* 54166 54265: gap of 100 bp  
 \* 54266 55074: contig of 809 bp in length  
 \* 55075 55174: gap of 100 bp  
 \* 55175 55952: contig of 778 bp in length  
 \* 55953 56052: gap of 100 bp  
 \* 56053 56825: contig of 773 bp in length  
 \* 56826 56925: gap of 100 bp  
 \* 56926 57700: contig of 775 bp in length  
 \* 57701 57800: gap of 100 bp  
 \* 57801 58594: contig of 794 bp in length  
 \* 58595 58694: gap of 100 bp  
 \* 58695 59458: contig of 764 bp in length  
 \* 59459 59558: gap of 100 bp  
 \* 59559 60353: contig of 795 bp in length  
 \* 60354 60453: gap of 100 bp  
 \* 60454 61235: contig of 782 bp in length  
 \* 61236 61335: gap of 100 bp

Query Match 59.0%; Score 24.2; DB 2; Length 72012;  
 Best Local Similarity 72.5%; Pred. No. 82;  
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CAAACCTAGCTCAAGGTCATCAAACTAGGTCARAGGTC 40  
 Db 31476 CAAACCTAGCTCAAGGTCATCAAACTAGGTCARAGGTC 31437

RESULT 15  
 SPDEXCAP 14056 bp DNA linear BCT 07-JUL-2002  
 LOCUS S.pneumoniae dexB, cap3A, cap3B and cap3C genes and orfs.  
 DEFINITION 247210  
 ACCESSION 247210  
 VERSION 247210.1 GI:1658316  
 KEYWORDS alpha,1-6-glucosidase; cap3A gene; cap3B gene; cap3C gene; celluburonic acid synthase; dexB gene; glucosephosphate uridylyltransferase; uridine diphosphate glucose dehydrogenase.

SOURCE Streptococcus pneumoniae.  
 ORGANISM Streptococcus pneumoniae; Streptococcaceae; Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.  
 REFERENCE 1 (bases 5911 to 7095)  
 AUTHORS Arrecubieta, C., Lopez, R. and Garcia, E.  
 TITLE Molecular characterization of cap3A, a gene from the operon required for the synthesis of the capsule of Streptococcus pneumoniae type 3: sequencing of mutations responsible for the unencapsulated phenotype and localization of the capsular cluster on the pneumococcal chromosome  
 J. Bacteriol. 176 (20), 6375-6383 (1994)  
 J95014083  
 MEDLINE 7929009  
 PUBMED 7929009  
 REFERENCE 2 (bases 1 to 9704)  
 AUTHORS Arrecubieta, C., Garcia, E. and Lopez, R.  
 TITLE Sequence and transcriptional analysis of a DNA region involved in

the production of capsular polysaccharide in Streptococcus pneumoniae type 3  
 Gene 167 (1-2), 1-7 (1995)  
 JOURNAL 96144241  
 MEDLINE 8566758  
 PUBMED 8566758  
 REFERENCE 3 (bases 1 to 14056)  
 AUTHORS Garcia, E., Arrecubieta, C., Munoz, R., Mollerach, M. and Lopez, R.  
 TITLE A functional analysis of the Streptococcus pneumoniae genes involved in the synthesis of type 1 and type 3 capsular polysaccharides  
 Unpublished  
 JOURNAL 4  
 REFERENCE Garcia, E.  
 AUTHORS Direct Submission  
 TITLE Submitted (04-JAN-1995) Garcia E., Centro de Investigaciones Biologicas. CSIC., Microbiologia Molecular, Velazquez, 144, Madrid, Madrid, SPAIN, 28006  
 JOURNAL Revised by [5]  
 REFERENCE 5 (bases 1 to 14056)  
 AUTHORS Garcia, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-SEP-1996) Garcia E., Centro de Investigaciones Biologicas. CSIC., Microbiologia Molecular, Velazquez, 144, Madrid, Madrid, SPAIN, 28006  
 COMMENT On Nov 5, 1996 this sequence version replaced gi:695741.  
 FEATURES  
 source  
 1..14056  
 /organism="Streptococcus pneumoniae"  
 /strain="406"  
 /db\_xref="taxon:1313"  
 gene  
 1..697  
 /gene="dexB"  
 CDS  
 1..697  
 /gene="dexB"  
 /function="degradation of glucans"  
 /standard\_name="glucosidase"  
 /citation=[3]  
 /codon\_start=2  
 /transl\_table=11  
 /product="alpha, 1-6-glucosidase"  
 /protein\_id="CAA87400.1"  
 /db\_xref="GI:728717"  
 /db\_xref="SWISS-PROT:Q54796"  
 /translation="NSLFWNNHDLPRIVSIWGNQDVEYREKSAKAFALLHLMRGTPYI YQGEIGMTVPFETLQDVEISLNPREALKGVPIEIMDSIRVIGENMARTPMQ WDESKAGFSGQPLAVNNYEMINYOALANPDSIFTYOKLVQIRKENSMLRAD FELLDADKVPAYIRKGDERRFLVVANLSNEEQDLTVEGKVKSVLIENTLAQEVFEKQ ILVPWDAPCVBL"  
 terminator  
 1042..1087  
 /citation=[3]  
 /citation=[2]  
 1658..1689  
 /citation=[2]  
 /function="recognition sequence for the insertion sequence IS1202"  
 misc\_feature  
 3639..3643  
 /citation=[2]  
 3649..4341  
 /gene="orf1"  
 3649..4341  
 /gene="orf1"  
 /note="The deduced amino acid sequence contains a sugar transport signature (LIGFLAGGIGTSVILLLELDTHVK) typical of glucose transporters"  
 /citation=[3]  
 /citation=[2]  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CAA87401.1"  
 /db\_xref="GI:695742"  
 /db\_xref="SPTREMBL:Q54797"

```

/translation="MKEQNTLEIDVQLPRLWKRKLVLVLAITSSVAFAYSTFVI
KPEFTSTRIYVNRQKSGITNQDQAGSYLVKDYREIILSQVLEEVISDLKLD
LTKGLANKIVPVDTIRIVSVNDRVPEEASRIANSREVAQAQKIISITRVSQVA
TLSEARPAISPPSPPIKRNTLIGFLAGGIGTGVIVLLELDTVHKRPEDIEDTLQMT
LLGVNPLGKLGK"
RBS
4340..4344
/citations=[2]
4351..4815
/gene="orf2"
4351..4815
/gene="orf2"
/citations=[3]
/citations=[2]
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAA87402.1"
/db_xref="GI:695743"
/db_xref="SPTRMBL:054798"
/translation="MPTLEAOKKLFPIKAREVYNALCTNIQLSGDKLKVISVTSVN
PGEKTTYSNIAISPARAGYKTLIDGTRNSVMSGPPKSKREKTELEPLSGTADL
SHGLUCNTINLNFVQSGVSPNFTALQSKNFVVMVKIFSKEVSILSNWKR"
terminator
4851..4903
/citations=[3]
/citations=[2]
join(5819..5846,5853..7095)
/gene="cap3A"
5819..5824
/gene="cap3A"
/citations=[3]
/citations=[1]
/citations=[2]
/evidence=experimental
5841..5846
/gene="cap3A"
/citations=[3]
/citations=[1]
/citations=[2]
/evidence=experimental
5853..7095
/gene="cap3A"
/citations=[3]
/citations=[2]
/evidence=experimental
5893..5900
/gene="cap3A"
/citations=[2]
5911..7095
/gene="cap3A"
/function="synthesis of UDP-glucuronic acid from
UDP-glucose"
/standard_name="UDP-Glc dehydrogenase"
/note="The Cap3A protein has been biochemically
characterized, J. Bacteriol. 178:2971-2974(1996)"
/citations=[3]
/citations=[1]
/citations=[2]
/codon_start=1
/evidence=experimental
/transl_table=11
/product="uridine diphosphate glucose dehydrogenase"
/protein_id="CAA87403.1"
/db_xref="GI:1658317"
/db_xref="SWISS-PROT:Q57346"
/translation="MKIAGSGYGLSVLVAQHEVKVIDVKRVESINNRRKP
IKDAIEKYLEKELEASLDFAHYKDYVEYAIATPTNDVDLNQFDTSSVEAAIK
TCMEYNDCTIVIKSTTPEGVTKEVREKNTDRIIPSEFLRESKALYDNLPSRIVV
GTDLDDSELTKRQPADLLKGGAIKEEVPILVAENEAFVAKLFNTYLATRVYEN
BIDTYSVKGLNPNTIIDIYCDPRIGSYNNPSFGYGCYLPKDTQOLKASFRDVE
NLITAVQSNTRKRIATAGAILAQPSVGVIYRLIMKSDSDNFRSSAVAGVWERLDNY
RE"
join(7094..7099,7111..8364)

```

```

/gene="cap3B"
7094..7099
/gene="cap3B"
/citation=[2]
7111..8364
/gene="cap3B"
/function="Synthesis of type 3 pneumococcal capsule"
/note="Expression and characterization of this gene has
been recently reported, J. Exp. Med. 184:449-455(1996)"
/citation=[3]
/citation=[2]
/codon_start=1
/evidence=experimental
/transl_table=11
/product="Synthase of the type 3 pneumococcal capsular
polysaccharide"
/protein_id="CAA87404.1"
/db_xref="GI:1658318"
/db_xref="SPTRMBL:P72520"
/translation="MYTRFLMLDFFQNHDPHFFMLFFVILIRWAVIYFHAVRYKSY
SCSVDEKLFSSVILPVVDEPLNLFESVLRHSPSEIIVWINGPKNERLKLCHD
PREKLENNMTPICQYTPVPGKRNALLFGLSHVDQSQDITVLVDSQDVTWTPTLSELL
KPFVCDKIKGGVTRQKILDDPERNLVTMFANLLEIRAEAGTWKANSVTGKVCCLPGRY
IAFRTILRECIHEEFNMFHGFHKEVSDRSLNLTLLKGYKTVQMDTISVYVDAPT
SNKKFIRQOLRWAEQSYNNLKOTPMIRNAPLMFFIYFTOMILPMLLISVGVNIFLL
KIINITIVITASWBEIILYVLLGMIFFSGGGRNFRAMSRMKWYVYVFLIPVFIIVLSII
MCPIRLLGLMRCSDDLGMGTRNLTE"
join(8651..8654,8662..9582)
/gene="cap3C"
8651..8654

```

Query Match 58.5%; Score 24; DB 1; Length 14056;  
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
 Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AAACTAGGTCAAGGTCATCAAACTAGGTCAAGGTCA 41  
 DB 5519 AACACTATGGCAAAGGTCATCACTCAAGCTTAAAGTTAA 5480

Search completed: February 19, 2003, 23:00:39  
 Job time : 416.515 secs

gene

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic - nucleic search, using sw model

run on: February 19, 2003, 19:26:38 ; Search time 24.9187 Seconds  
(without alignments)  
3705.323 Million cell updates/sec

title: US-09-808-388-3  
perfect score: 41  
sequence: 1 caaaactaggtcaagggtca.....caaaactaggtcaagggtca 41

scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

searched: 2185239 seqs, 1125999159 residues

total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq.101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	41	100.0	41	22 AAI64305	PPAR response elem
2	41	100.0	332	22 AAI64308	Partial synthetic
3	27.8	67.8	67	24 ABA05563	Oligonucleotide #1
4	26.8	65.4	69	23 ABL58060	Human PPAR respons
5	25	61.0	38	22 AAI64304	PPAR response elem
6	22.6	55.1	63	24 ABL64300	PPAR dependent gen
7	22.4	54.6	3295	19 AAV41344	M. catarrhalis str
8	22.4	54.6	3349	19 AAV41341	M. catarrhalis str
9	22.4	54.6	7235	20 AAZ29910	Genomic DNA encodi

c	10	22	53.7	1092	21	AACT9002	Human secreted pro
	11	21.6	52.7	477	24	ABN71122	Streptococcus poly
	12	21.6	52.7	501	24	ABN71226	Streptococcus poly
	13	21.4	52.2	2595	22	AAH33385	Human colon cancer
	14	21.4	52.2	2930	24	ABQ88181	Human osteoblast d
	15	21.4	52.2	2930	24	ABK70285	Human lung cancer
	16	21.4	52.2	3044	24	ABQ88180	Human osteoblast d
	17	21.4	52.2	3044	24	ABK84552	Human cDNA differe
	18	21.4	52.2	3044	24	ABN97223	Gene #3721 used to
	19	21.4	52.2	3047	24	ABK09792	Human ovarian tumo
	20	21.4	52.2	3072	24	ABQ88182	Human osteoblast d
	21	21.4	52.2	3115	21	AAC98113	Human colon cancer
	22	21.2	51.7	532	22	ABA61819	Human foetal liver
	23	21.2	51.7	532	22	ABA29404	Probe #7870 for ge
	24	21.2	51.7	532	22	AAK10133	Human brain expres
	25	21.2	51.7	532	22	AAK56027	Human bone marrow
	26	21.2	51.7	532	22	AAI17209	Probe #142 for ge
	27	21.2	51.7	532	22	AAI41743	Probe #10429 used
	28	21.2	51.7	532	24	ABS10203	Human genome-deriv
	29	21	51.2	52	22	AAI64306	PPAR response elem
	30	21	51.2	476	24	ABQ56661	Human colon cancer
	31	21	51.2	675	22	AAK58774	Human immune/haema
	32	21	51.2	865	21	AAZ80044	Human colon cancer
	33	21	51.2	1144	21	AAA26674	Candida albicans p
	34	21	51.2	1502	23	ABL06809	Drosophila melanog
	35	21	51.2	3395	23	ABL12318	Drosophila melanog
	36	21	51.2	3502	23	ABL06808	Drosophila melanog
	37	21	51.2	5385	22	AAH18234	Human cDNA sequenc
	38	21	51.2	9218	22	AAK40054	Genomic sequence #
	39	21	51.2	9218	22	AAK82320	Human immune/haema
	40	21	51.2	9218	22	AAK91471	Human digestive sy
	41	21	51.2	325791	22	AAK95240	Human Oestrogen re
	42	21	51.2	1503900	22	AAK95240	Human neuregulin-1
	43	21	51.2	1503900	22	AAK96733	Human neuregulin-1
	44	20.8	50.7	8414	22	ABK27618	DNA encoding novel
	45	20.6	50.2	51345	22	AAI61373	Soybean 318013 reg

ALIGNMENTS

RESULT 1

AAI64305	ID	AAI64305 standard; DNA; 41 BP.
XX	AC	AAI64305;
XX	DT	15-NOV-2001 (first entry)
XX	DE	PPAR response element (DR1) 2 21.
XX	KW	PPAR response element; antiinflammatory; antiarthritic; cytostatic;
XX	KW	cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;
XX	KW	peroxisome proliferator activated receptor;
XX	KW	secreted non-pancreatic phospholipase A2; ss.
XX	OS	Synthetic.
XX	FN	WO200168845-A2.
XX	PD	20-SEP-2001.
XX	PF	14-MAR-2001; 2001WO-FR00759.
XX	PR	14-MAR-2000; 2000FR-000362.
XX	PA	13-APR-2000; 2000US-0196959.
XX	XX	(AVET ) AVENTIS PHARMA SA.
XX	XX	Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;
XX	XX	WPI; 2001-582451/65.







XX Peroxisome proliferator activated receptor (PPAR) dependent gene  
 PT transcription activators -  
 XX  
 XX Example 1; Page 4; /pp; Japanese.  
 XX The invention relates to natural products derived PPAR (peroxisome  
 CC proliferator activated receptor) dependent gene transcription activators.  
 CC The activity of PPAR dependent gene transcription activators may be  
 CC described as, anorectic, antidiabetic, hypoglycaemic and antilipemic.  
 CC PPAR dependent gene transcription activators are involved in the  
 CC activation of lipid metabolism. The products of the invention are used in  
 CC the prevention and treatment of obesity, diabetes, hyperglycaemia,  
 CC hyperlipidaemia and insulin resistance. The products are safe agents for  
 CC PPAR dependent gene activators. The current sequence represents a PPAR  
 CC dependent gene transcription activator related DNA sequence.  
 XX  
 SQ Sequence 63 BP; 9 A; 18 C; 15 G; 21 T; 0 other;  
 Query Match 55.1%; Score 22.6; DB 24; Length 63;  
 Best Local Similarity 75.7%; Pred. No. 14;  
 Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 5 ACTAGTCAAGGTCATCAAACTAGTCAAGGTC 41  
 |||||  
 DB 41 ACCAGGACAAAGGTCAGTTGACCGGACAAAGGTC 5  
 |||||  
 RESULT 7  
 AAV41344  
 ID AAV41344 standard; DNA; 3295 BP.  
 XX  
 AC AAV41344;  
 XX  
 DT 07-OCT-1998 (first entry)  
 XX  
 DE M. catarrhalis strain O46E UspA2 antigen encoding DNA.  
 XX  
 KW Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;  
 KW vaccine; otitis media; sinusitis; lower respiratory tract infection;  
 KW immunity enhancer; immunoassay reagent; ds.  
 XX  
 OS Moraxella catarrhalis.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1212..3086  
 FT /\*tag= a  
 FT /product= "UspA2 antigen of strain O46E"  
 XX  
 PN WO9828333-A2.  
 XX  
 PD 02-JUL-1998.  
 XX  
 PF 19-DEC-1997; 97WO-US23930.  
 XX  
 PR 20-DEC-1996; 96US-0033598.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen EJ;  
 PI Maciver I;  
 XX  
 DR WPI; 1998-377595/32.  
 DR P-PSDB; AAW68204.  
 XX  
 XX New peptide(s) containing the core epitope of Moraxella catarrhalis  
 PT Usp proteins - useful in, e.g. vaccines to prevent or treat M.  
 PT catarrhalis infection, and antibodies for passive immunisation  
 XX  
 PS Claim 29; Pages 150-152; 237pp; English.  
 XX  
 CC This DNA encodes a UspA2 antigen of Moraxella catarrhalis strain O46E.  
 CC Nucleic acid sequences encoding the UspA1 and A2 antigens of

CC M. catarrhalis isolates O35E, O46E, TTA24 and TTA37 can be used in  
 CC genetic vaccination. An antigenic composition or vaccine containing  
 CC antigenic peptides from UspA1 or UspA2 antigens are used to induce an  
 CC immune response in mammals against M. catarrhalis and can be used to  
 CC treat infections such as otitis media, sinusitis, lower respiratory  
 CC tract infections. They can also be used as immunity enhancers for other  
 CC bacterial, parasitic or viral antigens, to raise antibodies and as  
 CC immunoassay reagents for detecting specific antibodies. The antibodies  
 CC are useful for passive immunisation and as immunoassay reagents.  
 CC Detection of the epitopic core sequence, by immunoassay or by PCR, is  
 CC used to diagnose infection. The Usp antigens encoding nucleic acid  
 CC sequences are also used to produce recombinant proteins and for screening  
 CC for potential anti-M. catarrhalis agents, while their fragments are  
 CC useful as diagnostic probes or primers or to isolate variant sequences.  
 XX  
 SQ Sequence 3295 BP; 1102 A; 609 C; 684 G; 900 T; 0 other;  
 Query Match 54.6%; Score 22.4; DB 19; Length 3295;  
 Best Local Similarity 72.5%; Pred. No. 34;  
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 CAAACTAGGTCAAAGGTCATCAAACTAGTCAAGGTC 40  
 |||||  
 DB 2118 CAAGTATATCAAGATCTTCAGAGGAGGTCGAAGGTC 2157  
 |||||  
 RESULT 8  
 AAV41341  
 ID AAV41341 standard; DNA; 3349 BP.  
 XX  
 AC AAV41341;  
 XX  
 DT 07-OCT-1998 (first entry)  
 XX  
 DE M. catarrhalis strain O35E UspA1 antigen encoding DNA.  
 XX  
 KW Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;  
 KW vaccine; otitis media; sinusitis; lower respiratory tract infection;  
 KW immunity enhancer; immunoassay reagent; ds.  
 XX  
 OS Moraxella catarrhalis.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 321..2816  
 FT /\*tag= a  
 FT /product= "UspA1 antigen of strain O35E"  
 XX  
 PN WO9828333-A2.  
 XX  
 PD 02-JUL-1998.  
 XX  
 PF 19-DEC-1997; 97WO-US23930.  
 XX  
 PR 20-DEC-1996; 96US-0033598.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen EJ;  
 PI Maciver I;  
 XX  
 DR WPI; 1998-377595/32.  
 DR P-PSDB; AAW68201.  
 XX  
 XX New peptide(s) containing the core epitope of Moraxella catarrhalis  
 PT Usp proteins - useful in, e.g. vaccines to prevent or treat M.  
 PT catarrhalis infection, and antibodies for passive immunisation  
 XX  
 PS Claim 23; Pages 136-138; 237pp; English.  
 XX  
 CC This DNA encodes a UspA1 antigen of Moraxella catarrhalis strain O35E.  
 CC Nucleic acid sequences encoding the UspA1 and A2 antigens of  
 CC M. catarrhalis isolates O35E, O46E, TTA24 and TTA37 can be used in  
 CC genetic vaccination. An antigenic composition or vaccine containing

CC antigenic peptides from UspA1 or UspA2 antigens are used to induce an  
 CC immune response in mammals against M. catarrhalis and can be used to  
 CC treat infections such as otitis media, sinusitis, lower respiratory  
 CC tract infections. They can also be used as immunity enhancers for other  
 CC bacterial, parasitic or viral antigens, to raise antibodies and as  
 CC immunosassay reagents for detecting specific antibodies. The antibodies  
 CC are useful for passive immunisation and as immunoassay reagents.  
 CC Detection of the epitopic core sequence, by immunoassay or by PCR, is  
 CC used to diagnose infection. The Usp antigens encoding nucleic acid  
 CC sequences are also used to produce recombinant proteins and for screening  
 CC for potential anti-M. catarrhalis agents, while their fragments are  
 CC useful as diagnostic probes or primers or to isolate variant sequences.  
 CC  
 CC Sequence 3349 BP; 1157 A; 650 C; 744 G; 798 T; 0 other;

Query Match 54.6%; Score 22.4; DB 19; Length 3349;  
 Best Local Similarity 72.5%; Pred. No. 34;  
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

NY 1 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTC 40  
 |||||  
 b 1311 CAGATGATATCAAGATCTTCAGAGGAGTGAAGGTC 1350

RESULT 9  
 LAZ29910/c  
 ID AAZ29910 standard; DNA; 7235 BP.  
 AC AAZ29910;  
 CC  
 CC 26-JAN-2000 (first entry)  
 CC  
 CC Genomic DNA encoding a SC20 protein of soybean.  
 CC  
 CC Promoter; seed coat specificity; SC20 protein; biological control;  
 CC herbicide resistance gene; viral coat protein; biological control;  
 CC Bt toxin; seed taste; ss.

Glycine max.  
 Key Location/Qualifiers  
 Promoter 1..2450  
 /\*tag= a  
 /\*note= "claimed"

WO9953067-A2.  
 21-OCT-1999.  
 13-APR-1999; 99WO-CA00293.  
 13-APR-1998; 98US-0059090.  
 (MIAC ) CANADA MIN AGRIC & AGRI-FOOD CANADA.  
 Miki B, Gijzen M, Miller S, Bowman L, Batchelor A, Hu M;  
 Boutillier K;  
 WPI; 1999-611304/52.  
 Novel promoter sequences and genes useful for inducing expression of  
 genes in plant seed coats -  
 Claim 34; Page 139-146; 155pp; English.

The present sequence encodes a SC20 protein. The promoter of  
 this gene is differentially expressed in seed coat tissues,  
 specifically within the outer integument, the inner integument, the  
 thick walled parenchyma, the thin walled parenchyma, the endodermium,  
 the hourglass cells, the palisade, the stellate parenchyma, or the  
 membranous endocarp associated with the seed coat. The seed-coat  
 promoters may be used to express proteins of interest in seed coat  
 tissues. Genes of interest include but are not restricted to herbicide

CC resistance genes, genes encoding viral coat proteins, or genes encoding  
 CC proteins conferring biological control of pests or pathogens, e.g. a Bt  
 CC toxin. Other genes that may be expressed include proteins that alter the  
 CC taste of the seed and/or affect the nutritive value of the seed.  
 CC  
 CC Sequence 7235 BP; 2430 A; 1177 C; 1218 G; 2410 T; 0 other;

Query Match 54.6%; Score 22.4; DB 20; Length 7235;  
 Best Local Similarity 72.5%; Pred. No. 40;  
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 1 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTC 40  
 |||||  
 Db 6230 CAAACTAGGTCATAGCTTACCAACAAAGTCGAACCTC 6191

RESULT 10  
 AAC79002/c  
 ID AAC79002 standard; DNA; 1092 BP.  
 XX AAC79002;  
 AC AAC79002;  
 CC  
 CC 14-FEB-2001 (first entry)  
 CC  
 CC Human secreted protein gene 6 clone HSJAN83.  
 CC  
 CC Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 CC anti-allergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 CC vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 CC cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 CC neurological disease; infection; human; secreted protein; ss.

OS Homo sapiens.  
 XX  
 XX WO200058358-A1.

PD 05-OCT-2000.  
 XX  
 XX 23-MAR-2000; 2000WO-US07725.  
 XX  
 XX 26-MAR-1999; 99US-0126602.  
 PR 14-JAN-2000; 2000US-0176063.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;  
 WPI; 2000-594640/56.  
 DR P-FSDB; AAB44340.

XX Forty nine nucleic acid molecules encoding human secreted proteins,  
 PT useful in the prevention, treatment and diagnosis of cancer, immune  
 PT disorders, cardiovascular disorders and neurological diseases -  
 XX  
 XX Claim 1; Page 318; 367pp; English.

XX The invention relates to the isolation of genes AAC78997-C79045 encoding  
 CC 49 human secreted proteins AAB4335-B44382. The genes can be used to  
 CC generate fusion proteins by linking to the gene for the human  
 CC immunoglobulin G Fc portion (AAC78988) for increasing the stability of  
 CC the fusion protein as compared to the human protein only. The genes and  
 CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated  
 CC from a range of human tissues disclosed in the specification. The  
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)  
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and

CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 SQ Sequence 1092 BP; 310 A; 224 C; 264 G; 280 T; 14 other;  
 Query Match 53.7%; Score 22; DB 21; Length 1092;  
 Best Local Similarity 73.7%; Pred. No. 40;  
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 QY 4 AACTAGGTCAAGGTGTCATCAAACTAGGTCAAGGTCA 41  
 |||||  
 DB 406 AACGAGCCAGTGTCTATGTAATTAAGTCAATTGTCA 369  
 |||||  
 RESULT 11  
 ABN71122  
 ID ABN71122 standard; DNA; 477 BP.  
 AC  
 XX  
 AC ABN71122;  
 XX  
 DT 01-JUL-2002 (first entry)  
 DE Streptococcus polynucleotide SEQ ID NO 10157.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200234771-A2.  
 XX  
 FN  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB04789.  
 XX  
 PR 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 DR WPI; 2002-352536/38.  
 DR P-PSDB; ABP30491.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX  
 PS Claim 7; Page 4145; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.

SQ Sequence 477 BP; 189 A; 78 C; 99 G; 111 T; 0 other;  
 Query Match 52.7%; Score 21.6; DB 24; Length 477;  
 Best Local Similarity 75.0%; Pred. No. 49;  
 Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 3 AAAGTGGTCAAGGTGTCATCAAACTAGGTCAAGG 38  
 |||||  
 DB 86 AAAGTGGTCAAGGTGTCATCAAACTAGGTCAAGG 121  
 |||||  
 RESULT 12  
 ABN67276  
 ID ABN67276 standard; DNA; 501 BP.  
 XX  
 AC ABN67276;  
 XX  
 DT 01-JUL-2002 (first entry)  
 DE Streptococcus polynucleotide SEQ ID NO 2465.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200234771-A2.  
 XX  
 FN  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB04789.  
 XX  
 PR 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 DR WPI; 2002-352536/38.  
 DR P-PSDB; ABP26645.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX  
 PS Claim 7; Page 3396; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.

Query Match 52.7%; Score 21.6; DB 24; Length 501;

Best Local Similarity 75.0%; Pred. No. 50;  
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

DY 3 AAACCTAGGTCAAGGTGATCAAACTAGGTCAAGG 38  
DB 110 AAACCTAGGTCAAGGTGATCAAACTAGGTCAAG 145

RESULT 13  
AAH33385/c  
ID AAH33385 standard; cDNA; 2595 BP.  
AC AAH33385;  
KT 03-SEP-2001 (first entry)  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:441.  
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
GW colorectal carcinoma; ss.  
KS Homo sapiens.  
NN WO200122920-A2.  
NX 05-APR-2001.  
PY 28-SEP-2000; 200WO-US26524.  
PZ 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
PX (HUMA-) HUMAN GENOME SCI INC.  
PY Ruben SM, Barash SC, Birse CE, Rosen CA;  
PI WPI; 2001-235357/24.  
DR P-PSDB; AAG73954.  
KX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
PS Claim 1; Page 2539-2540; 9803pp; English.  
KX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
cancer-associated nucleic acid molecules (N) and proteins (P), where  
the proteins are collectively known as colon cancer antigens. The colon  
cancer antigens have cytostatic activity and can be used in gene  
therapy and vaccine production. N and P may be used in the prevention,  
diagnosis and treatment of diseases associated with inappropriate P  
expression. For example, N and P may be used to treat disorders  
associated with decreased expression by rectifying mutations or deletions  
in a patient's genome that affect the activity of P by expressing  
inactive proteins or to supplement the patient's own production of P.  
Additionally, N may be used to produce the colon cancer-associated P,  
by inserting the nucleic acids into a host cell and culturing the cell  
to express the proteins. N and P can be used in the prevention, diagnosis  
and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
and AAB7789 represent sequences used in the exemplification of the  
present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
KX Sequence 2595 BP; 742 A; 562 C; 714 G; 567 T; 10 other;  
SQ

Query Match 52.2%; Score 21.4; DB 22; Length 2595;  
Best Local Similarity 71.8%; Pred. No. 80;  
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

DY 2 AAACCTAGGTCAAGGTGATCAAACTAGGTCAAGGTC 40  
DB 309 AATATTCATCCAGGGCATCACTCCAGCCAAAGGTC 271

RESULT 14  
ABQ88181/c  
ID ABQ88181 standard; cDNA; 2930 BP.  
XX  
AC ABQ88181;  
KT 18-SEP-2002 (first entry)  
DE Human osteoblast differentiation related cDNA SEQ ID NO 88.  
KW Human; osteoblast; stem cell differentiation; bone tissue deposition;  
GW osteoporosis; osteopathic; ss.  
KS Homo sapiens.  
NN WO200250301-A2.  
NX 27-JUN-2002.  
PY 18-DEC-2001; 2001WO-US48276.  
PZ 18-DEC-2000; 2000US-255882P.  
PR 24-APR-2001; 2001US-285691P.  
PX (GENE-) GENE LOGIC INC.  
PY (PROC) PROCTER & GAMBLE CO.  
PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;  
PI Mertz L;  
KX WPI; 2002-557663/59.  
KX Use of genes and their expression profiles associated with osteoblast  
PT differentiation for screening modulators bone formation, for diagnosing  
PT or treating e.g. osteoporosis, or as markers for the differentiation  
PT process -  
KX  
PS Claim 1; SEQ ID NO 88; 78pp + Sequence Listing; English.  
KX The invention relates to genes and their expression profiles are used  
CC for:  
CC (a) screening modulators of precursor stem cell differentiation into  
CC osteoblasts, or bone tissue deposition;  
CC (b) diagnosing abnormal deposition of bone tissue, abnormal rate of  
CC osteoblast formation or osteoporosis; or  
CC (c) treating or monitoring of bone tissue deposition.  
CC Monitoring the progression of bone tissue deposition.  
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid  
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy,  
CC drug-induced abnormalities in bone formation or bone loss, conditions  
CC that involve altered bone metabolism (e.g. idiopathic juvenile  
CC osteoporosis), skeletal disease linked to breast cancer, mastocytosis,  
CC Fanconi syndrome or fibrous dysplasia. The present sequence is that of an  
CC osteoblast differentiation associated cDNA marker of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
KX  
SQ Sequence 2930 BP; 793 A; 658 C; 821 G; 658 T; 0 other;  
Query Match 52.2%; Score 21.4; DB 24; Length 2930;  
Best Local Similarity 71.8%; Pred. No. 81;  
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

DY 2 AAACCTAGGTCAAGGTGATCAAACTAGGTCAAGGTC 40  
DB 757 AATATTCATCCAGGGCATCACTCCAGCCAAAGGTC 719

RESULT 15  
ABK70285/c

Search completed: February 19, 2003, 21:22:43  
Job time : 27.9187 secs

ID ABK70285 standard; cDNA; 2930 BP.  
XX AC ABK70285;  
XX DT 15-JUL-2002 (first entry)  
XX DE Human lung cancer associated full length cDNA DMSM-51.  
XX KW Human; ss; gene; lung cancer; cytostatic; tumour; vaccine.  
XX OS Homo sapiens.  
XX PN WO200224057-A2.  
XX PD 28-MAR-2002.  
XX PF 20-SEP-2001; 2001WO-US42232.  
XX PR 22-SEP-2000; 2000US-234837P.  
XX PR 10-OCT-2000; 2000US-239440P.  
XX PR 29-JUN-2001; 2001US-301928P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Benson DR, Mohamath R, Lodes MJ;  
XX DR WPI; 2002-372001/40.  
XX PT New tumour lung proteins and nucleic acids encoding the proteins, useful  
XX PT as vaccines and for treating, preventing, diagnosing or monitoring lung  
XX PT cancer  
XX PS Claim 1; Page 159-160; 189pp; English.  
XX  
XX The invention relates to an isolated polynucleotide comprising a sequence  
XX selected from 183 human DNA sequences (appearing as ABK70130-ABK70312),  
XX or their fragments, homologues, variants or complements and their encoded  
XX polypeptides. Also included are an expression vector comprising the  
XX polynucleotide operably linked to an expression control sequence; a host  
XX cell transformed or transfected with an expression vector of; an isolated  
XX antibody, or its antigen-binding fragment that specifically binds to the  
XX polypeptide; a method for detecting the presence of a cancer in a  
XX patient; a fusion protein comprising at least the polypeptide; an  
XX oligonucleotide that hybridises to the polynucleotide under moderately  
XX stringent conditions; a method for stimulating and/or expanding T cells  
XX specific for a tumour protein; an isolated T cell population comprising T  
XX cells prepared from the method of above; a composition comprising a first  
XX component consisting of carriers and immunostimulants, and a second  
XX component selected from the polynucleotides, proteins, antibodies, fusion  
XX proteins, T cell populations and antigen presenting cells expressing the  
XX polypeptide; methods for stimulating an immune response or treating  
XX cancer in a patient by administering the composition and diagnostic kits  
XX comprising at least one of the oligonucleotide of, or an antibody and a  
XX detection reagent consisting of a reporter group. The polypeptides and  
XX polynucleotides are useful as vaccines for the treatment or prevention of  
XX lung cancer, and for diagnosis and monitoring of such cancer. The  
XX polynucleotide, polypeptide and antigen presenting cells can be  
XX used to stimulate or expand T cells specific for a tumorous protein.  
XX The polynucleotides may be used as probes or primers for nucleic acid  
XX hybridisation, and in the preparation of ribozyme molecules for  
XX inhibiting expression of tumour polypeptides and proteins in tumour  
XX cells. The present sequence is one of the 183 lung cancer associated  
XX polynucleotides.  
XX  
SQ Sequence 2930 BP; 793 A; 658 C; 821 G; 658 T; 0 other;  
  
Query Match 52.2%; Score 21.4; DB 24; Length 2930;  
Best Local Similarity 71.8%; Pred. No. 81;  
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
  
QY 2 AAAACTAGGTCAAAGTCATCAAACTAGGTCAAAGGTC 40  
Db 757 ATATTTCAGTCCAGGGCATCACTCCAGGCCAAGGTC 719

; SEQ ID NO 2



```
; LENGTH: 3349
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-2

Query Match          54.6%; Score 22.4; DB 4; Length 3349;
Best Local Similarity 72.5%; Pred. No. 3.4;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAAGGTCTATCAAACTAGGTCAAAGGTC 40
    |||||
Db 1311 CAGATGATATCAAGATCTTTCAGAGGAGGTGAAAGGTC 1350

RESULT 3
US-09-328-111-128/c
; Sequence 128, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1998-06-08
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(865)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-128

Query Match          51.2%; Score 21; DB 4; Length 865;
Best Local Similarity 71.1%; Pred. No. 8.7;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAAGGTCTATCAAACTAGGTCAAAGG 38
    |||||
Db 595 CAAAATAGCAAGGCGCAAGGAACCTTTGCCAAGG 558

RESULT 4
US-09-328-111-129
; Sequence 129, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1998-06-08
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 129
; LENGTH: 910
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(910)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-129

Query Match          49.8%; Score 20.4; DB 4; Length 910;
Best Local Similarity 71.1%; Pred. No. 15;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAAGGTCTATCAAACTAGGTCAAAGG 38
    |||||
Db 154 CAAAATAGCAAGGCGCAAGGAACCTTTGCCAAGG 191

RESULT 5
US-08-123-343A-6/c
; Sequence 6, Application US/08123343A
; Patent No. 5593879
; GENERAL INFORMATION:
; APPLICANT: Steller, Hermann
; APPLICANT: Abrams, John M.
; APPLICANT: Grether, Megan E.
; APPLICANT: White, Kristin
; TITLE OF INVENTION: Cell Death Genes of Drosophila
; TITLE OF INVENTION: Melanogaster and Vertebrate Analogs
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,343A
; FILING DATE: 17-SEP-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/004,957
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-5907A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 61861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3900 base pairs
; TYPE: nucleic acid
```

```

; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-123-343A-6

Query Match 49.8%; Score 20.4; DB 1; Length 3900;
Best Local Similarity 71.1%; Pred. No. 22; Mismatches 0; Gaps 0;
Matches 27; Conservative 0; Indels 11;

2Y 2 AAAGTCTGCTCAAGGTCATCAAACTAGGTCAAGGT 39
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2136 AAATGCAATCAAGATTATCAAAAGTATCTCAAAAGT 2089

RESULT 6
US-07-686-591-3
; Sequence 3, Application US/07686591
; Patent No. 5215915
; GENERAL INFORMATION:
; APPLICANT: Tiberi, Mario
; APPLICANT: Jarvie, Keith R.
; APPLICANT: Caron, Marc G.
; TITLE OF INVENTION: Cloned Gene Encoding Rat D1B Dopamine Receptor
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and Gibson
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5215915th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/686,591
; FILING DATE: 19910406
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/686,591
; FILING DATE: 4/5/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2308 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 694..2118
; OTHER INFORMATION:
; US-07-686-591-3

Query Match 49.3%; Score 20.2; DB 1; Length 2308;
Best Local Similarity 68.3%; Pred. No. 23;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 CAAAGTCTGCTCAAGGTCATCAAACTAGGTCAAGGTCA 41
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1539 CAAGAGGAGACCAAGGTCCTCAAAACCCCTGTCAATGATCA 1579

RESULT 7
US-07-686-591-3
; Sequence 3, Application US/07686591
; Patent No. 5215915
; GENERAL INFORMATION:
; APPLICANT: Tiberi, Mario
; APPLICANT: Jarvie, Keith R.
; APPLICANT: Caron, Marc G.
; TITLE OF INVENTION: Cloned Gene Encoding Rat D1B Dopamine Receptor
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and Gibson
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5215915th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/686,591
; FILING DATE: 19910406
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2308 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 694..2118
; OTHER INFORMATION:
; US-07-686-591-3

Query Match 49.3%; Score 20.2; DB 1; Length 2308;
Best Local Similarity 68.3%; Pred. No. 23;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 CAAAGTCTGCTCAAGGTCATCAAACTAGGTCAAGGTCA 41
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1539 CAAGAGGAGACCAAGGTCCTCAAAACCCCTGTCAATGATCA 1579

RESULT 8
US-09-222-938A-47/C
; Sequence 47, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritze, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
```

NUMBER OF SEQ ID NOS: 102  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 47  
LENGTH: 1200  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (95)...(1126)  
US-09-222-938A-47

Query Match 48.8%; Score 20; DB 4; Length 1200;  
Best Local Similarity 72.2%; Pred. No. 24;  
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 6 CTAGGTCAAGGTCAATCAAACTAGGTCAAGGTCA 41  
Db 542 CTGTGCAAAATCCTTCAGAAATTTGTTCAAGGTTA 507

## RESULT 9

US-08-366-490-5  
Sequence 5, Application US/09366490  
Patent No. 5877403  
GENERAL INFORMATION:  
APPLICANT: McMaster, J. Russell  
APPLICANT: Boeshore, Maury L.  
APPLICANT: Tricoli, David M.  
APPLICANT: Reynolds, John F.  
APPLICANT: Carney, Kim J.  
TITLE OF INVENTION: PAPAYA RINGSPOT VIRUS PROTEASE GENE  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto  
STREET: 277 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10172-0194  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/366,490  
FILING DATE: 30-DEC-1994  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitzpatrick, Cella, Harper, and Scinto  
REFERENCE/DOCKET NUMBER: 4869  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-2400  
TELEFAX: 212-758-2982  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1797 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: PAPAYA RINGSPOT VIRUS  
STRAIN: P-TYPE  
INDIVIDUAL ISOLATE: USA (HA attenuated)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..1782  
FEATURE:  
NAME/KEY: mat\_peptide

LOCATION: 3..191  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 192..362  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 363..1643  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 1644..1782  
US-08-366-490-5

Query Match 48.3%; Score 19.8; DB 2; Length 1797;  
Best Local Similarity 69.2%; Pred. No. 31;  
Matches 27; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 CAAAACTAGGTCAAGGTCAATCAAACTAGGTCAAAAGGT 39  
Db 1214 CAGAGCTAGGTTTAGGCCATCAAGCTGGGACACAGGT 1252

## RESULT 10

US-08-860-483A-5  
Sequence 5, Application US/08860483A  
Patent No. 6046384  
GENERAL INFORMATION:  
APPLICANT: McMaster, J. R.  
APPLICANT: Boeshore, Maury L.  
APPLICANT: Tricoli, David M.  
APPLICANT: Reynolds, John F.  
APPLICANT: Carney, Kim J.  
APPLICANT: Slighton, Jerry L.  
APPLICANT: Gonsalves, Dennis  
TITLE OF INVENTION: Papaya Ringspot Virus N1a Protease Gene  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rockey, Milnamow & Katz  
STREET: 180 N. Statson Avenue, 2 Prudential Plaza,  
STREET: Suite 4700  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/860,483A  
FILING DATE: 26-JUN-1997  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Mueller, Lisa V.  
REGISTRATION NUMBER: 38,978  
REFERENCE/DOCKET NUMBER: SV93801P0091US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3126165400  
TELEFAX: 3126165460  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1797 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..1779  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1782..1797  
US-08-860-483A-5

Query Match 48.3%; Score 19.8; DB 3; Length 1797;  
Best Local Similarity 69.2%; Pred. No. 31;  
Matches 27; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Y 1 CAAAACTAGGTCAAAGGTCATCAAAACTAGGTCAAAGGT 39  
Db 1214 CAGAGCTAGGTTTAGGGCCATGAAGCTGGGACAAAGGT 1252

## RESULT 11

S-08-366-490-7

Sequence 7, Application US/08366490

Patent No. 5877403

## GENERAL INFORMATION:

APPLICANT: McMaster, J. Russell

APPLICANT: Boeshore, Maury L.

APPLICANT: Tricoli, David M.

APPLICANT: Reynolds, John F.

APPLICANT: Carney, Kim J.

TITLE OF INVENTION: PAPAYA RINGSPOT VIRUS PROTEASE GENE

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto

STREET: 277 Park Avenue

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10172-0194

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/366,490

APPLICATION NUMBER: US/08/366,490

FILING DATE: 30-DEC-1994

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Fitzpatrick, Cella, Harper, and Scinto

REFERENCE/DOCKET NUMBER: 4869

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-758-2400

TELEFAX: 212-758-2992

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1900 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

ORGANISM: PAPAYA RINGSPOT VIRUS

STRAIN: P-TYPE

INDIVIDUAL ISOLATE: Hawaii

FEATURE:

NAME/KEY: CDS

LOCATION: 3, 1900

PUBLICATION INFORMATION:

AUTHORS: Yeh, SD

AUTHORS: Jan, F

AUTHORS: Chiang, C

AUTHORS: Doong, T

AUTHORS: Chen, M

AUTHORS: Chung, P

AUTHORS: Bau, H

TITLE: Complete nucleotide sequence and genetic

TITLE: organization of papaya ringspot virus.

JOURNAL: J. Gen. Virol.

VOLUME: 73

PAGES: 2531-  
DATE: 1992  
US-08-366-490-7

Query Match 48.3%; Score 19.8; DB 2; Length 1900;  
Best Local Similarity 69.2%; Pred. No. 32;  
Matches 27; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 CAAAACTAGGTCAAAGGTCATCAAAACTAGGTCAAAGGT 39  
Db 1226 CAGAGCTAGGTTTAGGGCCATGAAGCTGGGACAAAGGT 1264

## RESULT 12

US-08-860-483A-8

Sequence 8, Application US/08860483A

Patent No. 6046384

## GENERAL INFORMATION:

APPLICANT: McMaster, J. R.

APPLICANT: Boeshore, Maury L.

APPLICANT: Tricoli, David M.

APPLICANT: Reynolds, John F.

APPLICANT: Carney, Kim J.

APPLICANT: Slighton, Jerry L.

APPLICANT: Gonsalves, Dennis

TITLE OF INVENTION: Papaya Ringspot Virus Nia Protease Gene

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Rockey, Milnanow &amp; Katz

STREET: 180 N. Stetson Avenue, 2 Prudential Plaza,

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/860,483A

FILING DATE: 26-JUN-1997

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Mueller, Lisa V.

REGISTRATION NUMBER: 38,978

REFERENCE/DOCKET NUMBER: SVS3801P0091US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 3126165400

TELEFAX: 3126165460

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 1900 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: DNA (genomic)

US-08-860-483A-8

Query Match 48.3%; Score 19.8; DB 3; Length 1900;  
Best Local Similarity 69.2%; Pred. No. 32;  
Matches 27; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 CAAAACTAGGTCAAAGGTCATCAAAACTAGGTCAAAGGT 39  
Db 1226 CAGAGCTAGGTTTAGGGCCATGAAGCTGGGACAAAGGT 1264

## RESULT 13

US-08-860-483A-9

Sequence 9, Application US/08860483A

Patent No. 6046384

GENERAL INFORMATION:  
APPLICANT: McMaster, J. R.  
APPLICANT: Boesroere, Maury L.  
APPLICANT: Tricoli, David M.  
APPLICANT: Reynolds, John F.  
APPLICANT: Carney, Kim J.  
APPLICANT: Slighton, Jerry L.  
APPLICANT: Gonsalves, Dennis  
TITLE OF INVENTION: Papaya Ringspot Virus Nia Protease Gene  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rockey, Milnamow & Katz  
STREET: 180 N. Stetson Avenue, 2 Prudential Plaza,  
SUITE: Suite 4700  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/860,483A  
FILING DATE: 26-JUN-1997  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Mueller, Lisa V.  
REGISTRATION NUMBER: 38,978  
REFERENCE/DOCKET NUMBER: SVS3801P0091US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3126165400  
TELEFAX: 3126165460  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1900 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: DNA (genomic)  
US-08-860-483A-9

Query Match 48.3%; Score 19.8; DB 3; Length 1900;  
Best Local Similarity 69.2%; Pred. No. 32;  
Matches 27; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 1 CAAACTAGGTCAAGGTCTATCAAACTAGGTCAAGGT 39  
Db 1226 CAGAGCTAGGTTTAGGCGCATGAAGCTGGGACAAAGT 1264

RESULT 14  
US-08-348-891A-1/c  
Sequence 1, Application US/08348891A  
Patent No. 5654136  
GENERAL INFORMATION:  
APPLICANT: SASAKI, Keiko  
APPLICANT: MORI, Takayuki  
APPLICANT: MAKINO, Satoshi  
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,  
CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
ITS ABSOLUTE IDENTIFICATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/348,891A  
FILING DATE: 25-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,400  
FILING DATE: 10-MAR-1992  
APPLICATION NUMBER: JP 3-293625  
FILING DATE: 14-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: KP-7501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297  
TELEFAX: 703-685-0573  
TELEX: 248425 EMEON  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15894 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 108..1682  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1807..3327  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3438..4442  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5458..7107  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 7271..9121  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9234..15782  
US-08-348-891A-1  
Query Match 48.3%; Score 19.8; DB 1; Length 15894;  
Best Local Similarity 77.4%; Pred. No. 54;  
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 11 TCAAGGTCTATCAAACTAGGTCAAGGTCA 41  
Db 9931 TCACATACATCAAAACACAGTTCAAAATGTCA 9901  
RESULT 15  
US-08-905-817-1/c  
Sequence 1, Application US/08905817  
Patent No. 5824777  
GENERAL INFORMATION:  
APPLICANT: SASAKI, Keiko  
APPLICANT: MORI, Takayuki  
APPLICANT: MAKINO, Satoshi  
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,  
CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
ITS ABSOLUTE IDENTIFICATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington

STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/905,817  
FILING DATE: 04-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,891  
FILING DATE: 25-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,400  
FILING DATE: 10-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 3-293625  
FILING DATE: 14-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: KP-7501A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297  
TELEFAX: 703-685-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15894 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 108..1682  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1807..3327  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3438..4442  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5458..7107  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 7271..9121  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9234..15782  
JS-08-905-817-1

Query Match 48.3%; Score 19.8; DB 1; Length 15894;  
Best Local Similarity 77.4%; Pred. No. 54;  
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 11 TCAGAGGTCATCAAACTAGGTCAAGGTCA 41  
DB 9931 TCACATACATCAAAACCAGTTCATATGCA 9901

Search completed: February 20, 2003, 01:13:45  
Job time : 12.2397 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:17:14 ; Search time 31.1001 Seconds  
(without alignments)  
671.453 Million cell updates/sec

Title: US-09-808-388-3  
Perfect score: 41  
Sequence: 1 caaaactaggtcaaaagtca.....caaaactaggtcaaaagtca 41

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/us07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	41	10	US-09-808-388-3
2	41	100.0	332	10	US-09-808-388-6
3	29	70.7	60	9	US-09-877-705A-142
4	29	70.7	60	9	US-09-877-738A-142
5	25	61.0	38	10	US-09-808-388-2
6	21.4	52.2	2930	10	US-09-960-253-156
7	21.4	52.2	3044	10	US-09-880-107-3718
8	21.4	52.2	3047	10	US-09-864-864-329
9	21.4	52.2	3115	10	US-09-925-299-123
10	21.2	51.7	532	10	US-09-864-761-7870
11	21	51.2	52	10	US-09-808-388-4
12	21	51.2	865	10	US-09-879-536-128
13	21	51.2	1503841	9	US-09-946-807-1
14	21	51.2	1503841	10	US-09-795-668-1
15	21	51.2	1503841	10	US-09-795-668-1
16	20.8	50.7	8414	9	US-09-764-868-1278
17	20.6	50.2	513509	9	US-09-754-853A-4
18	20.4	49.8	317	10	US-09-920-300A-819
19	20.4	49.8	317	10	US-09-998-598-2092

20	20.4	49.8	317	12	US-10-033-528-819	Sequence 819, App
C 21	20.4	49.8	595	9	US-09-796-692-8696	Sequence 8696, Ap
C 22	20.4	49.8	603	9	US-10-046-935-1043	Sequence 1043, Ap
C 23	20.4	49.8	603	9	US-09-878-178-1043	Sequence 1043, Ap
C 24	20.4	49.8	910	10	US-09-879-536-129	Sequence 129, App
C 25	20.4	49.8	3902	9	US-10-108-605-88	Sequence 88, Appl
C 26	20.2	49.3	1659	10	US-09-925-299-190	Sequence 190, App
C 27	20.2	49.3	5721	10	US-09-785-770A-15	Sequence 15, Appl
C 28	20.2	49.3	8121	10	US-09-785-770A-14	Sequence 14, Appl
C 29	20.2	49.3	14175	9	US-09-764-868-1474	Sequence 1474, Ap
C 30	20	48.8	20	9	US-09-877-705A-67	Sequence 67, Appl
C 31	20	48.8	20	9	US-09-877-705A-68	Sequence 68, Appl
C 32	20	48.8	20	9	US-09-877-738A-67	Sequence 67, Appl
C 33	20	48.8	20	9	US-09-877-738A-68	Sequence 68, Appl
C 34	20	48.8	20	10	US-09-808-388-1	Sequence 1, Appli
C 35	20	48.8	1400	10	US-09-887-576-526	Sequence 526, App
C 36	20	48.8	1400	10	US-09-887-576-537	Sequence 537, App
C 37	20	48.8	1523	9	US-09-938-842A-4708	Sequence 4708, Ap
C 38	20	48.8	2000	9	US-09-938-842A-4013	Sequence 4013, Ap
C 39	20	48.8	2003	10	US-09-887-576-8	Sequence 8, Appli
C 40	20	48.8	3697	12	US-10-002-600-58	Sequence 58, Appl
C 41	19.8	48.3	426	10	US-09-960-352-13817	Sequence 13817, A
C 42	19.8	48.3	484	10	US-09-864-761-4318	Sequence 4318, Ap
C 43	19.8	48.3	624	10	US-09-998-598-1265	Sequence 1265, Ap
C 44	19.8	48.3	649	10	US-09-747-155-392	Sequence 392, App
C 45	19.8	48.3	649	10	US-09-747-155-418	Sequence 418, App

ALIGNMENTS

RESULT 1

US-09-808-388-3  
; Sequence 3, Application US/09808388  
; Patent No. US20020081719A1  
; GENERAL INFORMATION: Charbel  
; APPLICANT: Massaad, Charbel  
; APPLICANT: Berenbaum, Francis  
; APPLICANT: Olivier, Jean-Luc  
; APPLICANT: Salvat, Colette  
; APPLICANT: Berezat, Gilbert  
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
; TITLE OF INVENTION: their uses  
; FILE REFERENCE: ST00010  
; CURRENT APPLICATION NUMBER: US/09/808,388  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: FR/00/03262  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: US 60/196,959  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 41  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: PPPE element  
US-09-808-388-3

Query Match 100.0%; Score 41; DB 10; Length 41;  
Best Local Similarity 100.0%; Pred. No. 1e-06;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAAACCTAGGTCAAAAGTCAATCAAGTCAAGTCA 41  
DB 1 CAAAACCTAGGTCAAAAGTCAATCAAGTCAAGTCA 41

RESULT 2

US-09-808-388-6  
; Sequence 6, Application US/09808388  
; Patent No. US20020081719A1

```

; GENERAL INFORMATION:
; APPLICANT: Massaad, Charbel
; APPLICANT: Berendaum, Francis
; APPLICANT: Olivier, Jean-Luc
; APPLICANT: Salvat, Colette
; APPLICANT: Berziat, Gilbert
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them
; FILE REFERENCE: ST00010
; CURRENT APPLICATION NUMBER: US/09/808,388
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: FR/00/03262
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: US 60/196,959
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 6
; LENGTH: 332
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PPPE/PLA2s hybrid promoter
US-09-808-388-6

Query Match 100.0%; Score 41; DB 10; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTCA 41
Db 13 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTCA 53

RESULT 3
US-09-877-705A-142/c
; Sequence 142, Application US/09877705A
; Publication No. US20030008282A1
; GENERAL INFORMATION:
; APPLICANT: Li, Jason
; TITLE OF INVENTION: METHOD FOR SCREENING FOR DRUG CANDIDATES FOR MODULATING TRANSCRIPTION ACTIVITY
; FILE REFERENCE: 26757-704
; CURRENT APPLICATION NUMBER: US/09/877,705A
; CURRENT FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 162
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 142
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hybridization probe MP68
US-09-877-705A-142

Query Match 70.7%; Score 29; DB 9; Length 60;
Best Local Similarity 97.6%; Pred. No. 0.029;
Matches 40; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTCA 41
Db 40 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTCA 1

RESULT 4
US-09-877-738A-142/c
; Sequence 142, Application US/09877738A
; Publication No. US20030022173A1
; GENERAL INFORMATION:
; APPLICANT: Li, Jason
; TITLE OF INVENTION: METHOD AND KIT FOR ISOLATING DNA PROBES THAT BIND TO ACTIVATED TRANSCRIPTION FACTORS
; FILE REFERENCE: 26757-701

```

```

; CURRENT APPLICATION NUMBER: US/09/877,738A
; CURRENT FILING DATE: 2001-06-01
; NUMBER OF SEQ ID NOS: 162
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 142
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hybridization probe MP68
US-09-877-738A-142

Query Match 70.7%; Score 29; DB 9; Length 60;
Best Local Similarity 97.6%; Pred. No. 0.029;
Matches 40; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTCA 41
Db 40 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTCA 1

RESULT 5
US-09-808-388-2
; Sequence 2, Application US/09808388
; Patent No. US20020081719A1
; GENERAL INFORMATION:
; APPLICANT: Massaad, Charbel
; APPLICANT: Berendaum, Francis
; APPLICANT: Olivier, Jean-Luc
; APPLICANT: Salvat, Colette
; APPLICANT: Berziat, Gilbert
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them
; FILE REFERENCE: ST00010
; CURRENT APPLICATION NUMBER: US/09/808,388
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: FR/00/03262
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: US 60/196,959
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 38
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PPPE element
US-09-808-388-2

Query Match 61.0%; Score 25; DB 10; Length 38;
Best Local Similarity 92.7%; Pred. No. 0.8;
Matches 38; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTCA 41
Db 1 CAAACTAGGTCAAAGGTCATCAAACTAGGTCAAAGGTCA 38

RESULT 6
US-09-960-253-156/c
; Sequence 156, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121-556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187

```



```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-960-253-156

Query Match      52.2%; Score 21.4; DB 10; Length 2930;
Best Local Similarity 71.8%; Pred. No. 40;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AAAAAGTGGTCAAAGGTCATCAAAAGTGGTCAAAGGTC 40
   |||||
Db 757 AATATTCAGTCCAAAGGTCATCACTCCAGCCAAAGGTC 719

RESULT 7
US-09-880-107-3718/c
; Sequence 3718, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3718
; LENGTH: 3044
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X51521
; NAME/KEY: unsure
; LOCATION: (1)..(3044)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3718

Query Match      52.2%; Score 21.4; DB 10; Length 3044;
Best Local Similarity 71.8%; Pred. No. 40;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AAAAAGTGGTCAAAGGTCATCAAAAGTGGTCAAAGGTC 40
   |||||
Db 798 AATATTCAGTCCAAAGGTCATCACTCCAGCCAAAGGTC 760

RESULT 8
US-09-864-864-329/c
; Sequence 329, Application US/09864864
; Patent No. US20020102679A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Secrist, Heather
; APPLICANT: Lodes, Michael J.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steve P.
; APPLICANT: Mannion, Jane
; APPLICANT: Benson, Darin R.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OTHER INFORMATION: AND DIAGNOSIS OF OVARIAN CANCER
US-09-864-864-329/c

; FILE REFERENCE: 210121.523
; CURRENT APPLICATION NUMBER: US/09/864,864
; CURRENT FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 341
; TYPE: DNA
; ORGANISM: Corixa Invention Disclosure Database
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 329
; LENGTH: 3047
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(3047)
; OTHER INFORMATION: n = A,T,C or G
US-09-864-864-329

Query Match      52.2%; Score 21.4; DB 10; Length 3047;
Best Local Similarity 71.8%; Pred. No. 40;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AAAAAGTGGTCAAAGGTCATCAAAAGTGGTCAAAGGTC 40
   |||||
Db 798 AATATTCAGTCCAAAGGTCATCACTCCAGCCAAAGGTC 760

RESULT 9
US-09-925-299-123/c
; Sequence 123, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 123
; LENGTH: 3115
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-09-925-299-123

Query Match      52.2%; Score 21.4; DB 10; Length 3115;
Best Local Similarity 71.8%; Pred. No. 40;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AAAAAGTGGTCAAAGGTCATCAAAAGTGGTCAAAGGTC 40
   |||||
Db 830 AATATTCAGTCCAAAGGTCATCACTCCAGCCAAAGGTC 792

RESULT 10
US-09-864-761-7870/c
; Sequence 7870, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; OTHER INFORMATION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aesmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
US-09-864-761-7870/c
```

PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 7870  
LENGTH: 532  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC015842.4  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1  
US-09-864-761-7870

Query Match 51.7%; Score 21.2; DB 10; Length 532;  
Best Local Similarity 76.5%; Pred. No. 34;  
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 TAGGTCAAGGTCATCAAACTAGGTCAAGGTC 40  
Db 75 TTGTTCAAGGTCACCTATAATAGTCAATGTC 42

RESULT 11  
US-09-808-388-4  
Sequence 4, Application US/09808388  
Patent No. US2002008179A1  
GENERAL INFORMATION:  
APPLICANT: Massead, Charbel  
APPLICANT: Berensbaum, Francis  
APPLICANT: Olivier, Jean-Luc  
APPLICANT: Salvat, Colette  
APPLICANT: Bereziat, Gilbert  
TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
FILE REFERENCE: ST00010  
CURRENT APPLICATION NUMBER: US/09/808,388

CURRENT FILING DATE: 2001-09-20  
PRIOR APPLICATION NUMBER: FR/00/03262  
PRIOR FILING DATE: 2000-03-14  
PRIOR APPLICATION NUMBER: US 60/196,959  
PRIOR FILING DATE: 2000-04-13  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 52  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: PPR element  
US-09-808-388-4

Query Match 51.2%; Score 21; DB 10; Length 52;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAAGGTCAT 21  
Db 1 CAAACTAGGTCAAGGTCAT 21

RESULT 12  
US-09-879-536-128/c  
Sequence 128, Application US/09879536  
Patent No. US20020144298A1  
GENERAL INFORMATION:  
APPLICANT: Endege, Wilson O.  
APPLICANT: Steinmann, Kathleen E.  
APPLICANT: Astle, Jon H.  
APPLICANT: Burgess, Christopher C.  
APPLICANT: Bushnell, Steven E.  
APPLICANT: Carroll III, Eddie  
APPLICANT: Catino, Theodore J.  
APPLICANT: Derti, Adnan  
APPLICANT: Ford, Donna M.  
APPLICANT: Lewis, Marcia E.  
APPLICANT: McNahan, John E.  
APPLICANT: Schlegel, Robert  
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
FILE REFERENCE: CCD-257 (US)  
CURRENT APPLICATION NUMBER: US/09/879,536  
CURRENT FILING DATE: 2001-09-21  
PRIOR APPLICATION NUMBER: US 60/088,801  
PRIOR FILING DATE: 1998-06-10  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 128  
LENGTH: 865  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(865)  
OTHER INFORMATION: n = A,T,C or G  
US-09-879-536-128

Query Match 51.2%; Score 21; DB 10; Length 865;  
Best Local Similarity 71.1%; Pred. No. 44;  
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAAGGTCATCAAACTAGGTCAAGG 38  
Db 595 CAAATAAGNCAAGGCGCAAGGAACCTTGCCAAAGG 558

RESULT 13  
US-09-946-807-1  
Sequence 1, Application US/09946807  
Patent No. US20020165144A1

GENERAL INFORMATION:  
APPLICANT: Stefansson, Hreinn  
APPLICANT: Steinthorsdottir, Valgerdur  
APPLICANT: Gulcher, Jeffrey R.  
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
FILE REFERENCE: 2345.2004-001  
CURRENT APPLICATION NUMBER: US/09/946,807  
CURRENT FILING DATE: 2001-09-05  
PRIOR APPLICATION NUMBER: US/09/795,668  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 09/515,716  
PRIOR FILING DATE: 2000-02-28  
NUMBER OF SEQ ID NOS: 1531  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1503841  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: r=g or a  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: y=t/u or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: m=a or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: k=g or t/u  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: s=g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: w=a or t/u  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: b=g or c or t/u  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: d=a or g or t/u  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: h=a or c or t/u  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: v=a or g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: n=a or g or c or t/u  
US-09-946-807-1

Query Match 51.2%; Score 21; DB 9; Length 1503841;  
Best Local Similarity 82.8%; Pred. No. 1.5e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 13 AAAGTCATCAAACTAGGTCAAGGTCA 41  
DB 226655 AAAAGTCTTCAAAAAAGGTCAAGGTCA 226683

RESULT 14  
US-09-795-668-1  
; Sequence 1, Application US/09795668  
; Patent No. US20020045577A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2004-001  
; CURRENT APPLICATION NUMBER: US/09/795,668  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,716  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1503841  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: r=g or a  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: y=t/u or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: m=a or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: k=g or t/u  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: s=g or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: w=a or t/u  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: b=g or c or t/u  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: d=a or g or t/u  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: h=a or c or t/u  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: v=a or g or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1531)  
; OTHER INFORMATION: n=a or g or c or t/u  
US-09-795-668-1

Query Match 51.2%; Score 21; DB 10; Length 1503841;  
Best Local Similarity 82.8%; Pred. No. 1.5e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 13 AAAGTCATCAAACTAGGTCAAGGTCA 41  
DB 226655 AAAAGTCTTCAAAAAAGGTCAAGGTCA 226683

RESULT 15  
US-09-795-686-1  
; Sequence 1, Application US/09795686  
; Patent No. US20020094954A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.

```

/ TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
/ FILE REFERENCE: 2345.2005-001
/ CURRENT APPLICATION NUMBER: US/09/795,686
/ CURRENT FILING DATE: 2001-02-28
/ PRIOR APPLICATION NUMBER: US 09/515,715
/ PRIOR FILING DATE: 2000-02-28
/ NUMBER OF SEQ ID NOS: 1531
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 1503841
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: x=g or a
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: y=t/u or c
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: m=a or c
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: k=g or t/u
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: s=g or c
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: w=a or t/u
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: b=g or c or t/u
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: d=a or g or t/u
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: h=a or c or t/u
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: v=a or g or c
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1531)
/ OTHER INFORMATION: n=a or g or c or t/u
/ US-09-795-686-1

```

```

Query Match      51.2%; Score 21; DB 10; Length 1503841;
Best Local Similarity 82.8%; Pred. No. 1.5e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 13 AAAGGTCTCATCAAACTAGGTCAAGGTCA 41
Db 226655 AAAGTCTTCAAAAAAGGTCCAGGTCA 226683

```

Search completed: February 20, 2003, 07:01:20  
Job time : 866.1 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:33:48 ; Search time 199.495 seconds  
(without alignments)  
3328.484 Million cell updates/sec

Title: US-09-808-388-3  
Perfect score: 41  
Sequence: 1 caaaataggtcaagggtca.....caaaactagggtcaagggtca 41

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

```

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hci:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_nam:*
24: em_gss_mus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26.6	64.9	112	17	FR0004396
2	25.6	62.4	374	17	FR0016627
3	25.6	62.4	636	14	BQ866576
4	25.6	62.4	686	14	BQ860634
5	25.6	62.4	697	14	BQ864107
6	25.6	62.4	740	14	BQ866858

7	25.6	62.4	771	14	BQ866155
8	24	58.5	442	17	FR0043044
9	24	58.5	709	14	BQ999579
10	23.8	58.0	237	9	AI780559
11	23.8	58.0	399	14	BQ505145
12	23.8	58.0	496	10	AW189153
13	23.8	58.0	500	10	AW189470
14	23.6	57.6	273	14	BQ868068
15	23.6	57.6	554	14	BQ862863
16	23.6	57.6	566	14	BQ996093
17	23.6	57.6	578	14	BQ866097
18	23.6	57.6	578	14	BQ859080
19	23.6	57.6	581	14	BQ860878
20	23.6	57.6	642	14	BU002480
21	23.6	57.6	644	14	BQ869839
22	23.6	57.6	651	14	BQ865773
23	23.6	57.6	657	14	BQ870317
24	23.6	57.6	682	14	BQ869250
25	23.6	57.6	688	14	BQ998669
26	23.6	57.6	689	14	BQ863909
27	23.6	57.6	710	14	BQ858443
28	23.6	57.6	714	14	BQ860933
29	23.6	57.6	714	14	BQ861217
30	23.6	57.6	714	14	BQ866965
31	23.6	57.6	743	14	BQ864826
32	23.6	57.6	743	14	BQ866827
33	23.6	57.6	750	14	BQ863352
34	23.6	57.6	761	14	BQ870563
35	23.4	57.1	308	17	FR0010836
36	23.4	57.1	454	14	BQ966442
37	23.4	57.1	484	14	BQ968221
38	23.4	57.1	510	14	BQ967096
39	23.4	57.1	544	10	AV844441
40	23.4	57.1	557	13	BM386002
41	23.4	57.1	612	17	FR0043020
42	23.2	56.6	132	9	AI151712
43	23.2	56.6	828	13	BF333289
44	23	56.1	294	12	BF349523
45	23	56.1	424	14	BQ965437

#### ALIGNMENTS

RESULT 1  
FR0004396 112 bp DNA linear GSS 27-FEB-1997  
LOCUS F.rubripes GSS sequence, clone 047P21aa2, genomic survey sequence.  
DEFINITION Z88179  
ACCESSION Z88179 GI:1885091  
VERSION Z88179.1  
KEYWORDS GSS; genome survey sequence.  
SOURCE Takifugu rubripes.  
ORGANISM Takifugu rubripes  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Takifugu.  
REFERENCE 1 (bases 1 to 112)  
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,  
Williams,G. and Brenner,S.  
TITLE Direct Submission  
JOURNAL Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource  
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmrc.mrc.ac.uk  
COMMENT Vector: ml3mp18  
V\_type: Phase  
PRIMER: M13  
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic  
sequence.  
FEATURES  
Location/Qualifiers  
1..112  
/organism="Takifugu rubripes"  
/db\_xref="taxon:31033"

BQ866155 QGC7D05.Y

AI130536 Fugu rubr

BQ999579 QGC22H18.

AI780559 EST261438

BQ505145 EST612560

AW189153 xl01h03.x

AW189470 xl07a10.x

BQ868068 QGD12P06.

BQ862863 QGC22F01.

BQ996093 QGC11N19.

BQ866097 QGC7A14.Y

BQ859080 QGC12A16.

BQ860878 QGC16O24.

BU002480 QGC31H16.

BQ869839 QGD7F20.Y

BQ865773 QGC5N11.Y

BQ870317 QGC8L12.Y

BQ869250 QGD5J18.Y

BQ998669 QGC1F11.Y

BQ863909 QGC25D05.

BQ858443 QGC10E05.

BQ860933 QGC17B08.

BQ861217 QGC17N16.

BQ866965 QGC9G20.Y

BQ864826 QGC27M10.

BQ866827 QGC9A20.Y

BQ863352 QGC23K10.

BQ870563 QGD9G11.Y

AI002107 F.rubripe

BQ966442 QHB26K01.

BQ968221 QHB33D21.

BQ967096 QHB28O09.

AV844441 AV844441

BM386002 UI-R-DN1-

AI130512 Fugu rubr

AI151712 SMCVAFCAP

BF333289 602996645

BF349523 MRC-HT015

BQ965437 QHB2IN07.

BASE COUNT 32 a 29 c 36 g 15 t  
 ORIGIN /clone="047P21a2"  
 /clone\_lib="cosmid 047P21"

Query Match 64.9%; Score 26.6; DB 17; Length 112;  
 Best Local Similarity 78.0%; Pred. No. 11;  
 Matches 32; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CAAGCTAGGTCAGGTCATCAAACTAGGTCAAGGTCA 41  
 |||||  
 Db 1 CAAGCTGGGACCAAGGTCACCAAGGCTGGACCAAGGTCA 41  
 |||||

## RESULT 2

FR0016617/c FR0016617 374 bp DNA linear GSS 19-SEP-1997  
 LOCUS F.rubripes GSS sequence, clone 187H11aF8, genomic survey sequence.  
 DEFINITION AL007814

ACCESSION AL007814.1 GI:2453384

VERSION GSS; genome survey sequence.

KEYWORDS Takifugu rubripes.

SOURCE Takifugu rubripes.

ORGANISM Takifugu rubripes

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 374)

Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umranta, Y.,

Williams, G. and Brenner, S.

Direct Submission

Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource

Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@gmp.mrc.ac.uk

Vector: pBluescript II KS

V type: phagemid

PRIMER: KS

DBSCR:

One pass dye-terminator sequencing of cosmid cloned genomic

sequence.

FEATURES Location/Qualifiers

1..374

/organism="Takifugu rubripes"

/db\_xref="taxon:31033"

/clone="187H11aF8"

/clone\_lib="cosmid 187H11"

BASE COUNT 75 a 112 c 89 g 93 t 5 others

## ORIGIN

Query Match 62.4%; Score 25.6; DB 17; Length 374;  
 Best Local Similarity 77.5%; Pred. No. 44;  
 Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 AAAGCTAGGTCAGGTCATCAAACTAGGTCAAGGTCA 41  
 |||||  
 Db 374 AAAGCTGGGACCAAGGTCACCAAGGCTGGGACCAAGGTCA 335  
 |||||

## RESULT 3

BQ866576 BQ866576 636 bp mRNA linear EST 14-AUG-2002  
 LOCUS BQ866576  
 DEFINITION QGC8F19.YG.ab1 QG-ABCDI lettuce salinas Lactuca sativa cDNA clone

QGC8F19, mRNA sequence.

ACCESSION BQ866576.1 GI:22252041

VERSION BQ866576

KEYWORDS EST.

SOURCE Lactuca sativa.

ORGANISM Lactuca sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;

Lactuca.

REFERENCE 1 (bases 1 to 636)

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,

Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.  
 Lettuce and Sunflower ESTs from the Compositae Genome Project  
 http://compgenomics.ucdavis.edu/  
 Unpublished (2002)

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]

belongs to contig QG\_CA\_Contig5488, see http://cgpdb.ucdavis.edu/

for details.

Plate: QGC8 row: F column: 19.

Location/Qualifiers

1..636

/organism="Lactuca sativa"

/cultivar="Salinas"

/db\_xref="taxon:4236"

/clone="QGC8F19"

/clone\_lib="QG-ABCDI lettuce salinas"

/lab\_host="E.Coli"

/note="Vector: pBRCNDSfiAB; The library was constructed

from 10 different sources of RNA from a single genotype.

Separate cDNAs were generated using primers that

incorporated unique 5' and 3' tags to distinguish each

source of RNA. cDNAs were then pooled, size-fractionated,

directionally cloned into a custom medium-copy vector and

transformations made with four size classes to minimize

size bias. Details of each source of RNA and library

construction can be obtained at http://cgpdb.ucdavis.edu/

TAG\_LIB-QG-ABCDI lettuce salinas

TAG\_TISSUE=chemical induction

TAG\_SEQ=GTGAGCCGGG

BASE COUNT 186 a 118 c 147 g 185 t

## ORIGIN

Query Match 62.4%; Score 25.6; DB 14; Length 636;  
 Best Local Similarity 87.5%; Pred. No. 57;  
 Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 AGGTCAAAGGTCATCAAACTAGGTCAAAGGT 39  
 |||||

Db 43 AGGTCAAAGGTCATCAAACTAGGTCAAAGGT 74  
 |||||

RESULT 4  
 BQ860634 BQ860634 686 bp mRNA linear EST 14-AUG-2002  
 LOCUS BQ860634  
 DEFINITION QGC15E12.YG.ab1 QG-ABCDI lettuce salinas Lactuca sativa cDNA clone

QGC15E12, mRNA sequence.

ACCESSION BQ860634

VERSION BQ860634.1 GI:22246099

KEYWORDS EST.

SOURCE Lactuca sativa.

ORGANISM Lactuca sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;

Lactuca.

REFERENCE 1 (bases 1 to 686)

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,

Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,

Church, S., Jackson, L. and Bradford, K.

Lettuce and Sunflower ESTs from the Compositae Genome Project

http://compgenomics.ucdavis.edu/

Unpublished (2002)

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozikeatgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig QG\_CA\_Contig5488, see http://cgpdb.ucdavis.edu/  
 for details.  
 Plate: QGC16 row: E column: 12.  
 Location/Qualifiers  
 1. .686  
 /organism="Lactuca sativa"  
 /cultivar="Salinas"  
 /db\_xref="taxon:4236"  
 /clone="QGC16E12"  
 /clone\_lib="QG ABCDI lettuce salinas"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/  
 TAG LIB=QG ABCDI lettuce salinas  
 TAG TISSUE=chemical induction  
 TAG\_SEQ=TGTAGCCGGG"  
 208 a 129 c 153 g 196 t  
 ORIGIN  
 Query Match 62.4%; Score 25.6; DB 14; Length 686;  
 Best Local Similarity 87.5%; Pred. No. 59;  
 Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 y 8 AGGTCAAAGGTCATCAAACTAGTCAAAAGGT 39  
 |||||  
 b 43 AGGTCAAAGGTCATCAAAATTTGGTCAAAAGTT 74  
 |||||  
 RESULT 5  
 Q0864107  
 LOCUS B0864107 .697 bp mRNA linear EST 14-AUG-2002  
 DEFINITION QGC25M20.yg.ab1 QG ABCDI lettuce salinas Lactuca sativa cDNA clone  
 QGC25M20, mRNA sequence.  
 ACCESION B0864107.1 GI:22249572  
 VERSION B0864107  
 KEYWORDS EST.  
 SOURCE Lactuca sativa.  
 ORGANISM Lactuca sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asterididae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 Lactuca.  
 1 (bases 1 to 697)  
 REFERENCE Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,  
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison  
 P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,  
 Church,S., Jackson,L. and Bradford,K.  
 Lettuce and Sunflower ESTs from the Compositae Genome Project  
 http://compgenomics.ucdavis.edu/  
 Unpublished (2002)  
 CONTACT: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozikeatgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig QG\_CA\_Contig5488, see http://cgpdb.ucdavis.edu/  
 for details.  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozikeatgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig QG\_CA\_Contig5488, see http://cgpdb.ucdavis.edu/  
 for details.  
 Plate: QGC25 row: M column: 20.  
 Location/Qualifiers  
 1. .697  
 /organism="Lactuca sativa"  
 /cultivar="Salinas"  
 /db\_xref="taxon:4236"  
 /clone="QGC25M20"  
 /clone\_lib="QG ABCDI lettuce salinas"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that

/organism="Lactuca sativa"  
 /cultivar="Salinas"  
 /db\_xref="taxon:4236"  
 /clone="QGC25M20"  
 /clone\_lib="QG ABCDI lettuce salinas"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/  
 TAG LIB=QG ABCDI lettuce salinas  
 TAG TISSUE=chemical induction  
 TAG\_SEQ=TGTAGCCGGG"  
 214 a 132 c 155 g 196 t  
 ORIGIN  
 Query Match 62.4%; Score 25.6; DB 14; Length 697;  
 Best Local Similarity 87.5%; Pred. No. 59;  
 Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 8 AGGTCAAAGGTCATCAAACTAGTCAAAAGGT 39  
 |||||  
 Db 43 AGGTCAAAGGTCATCAAAATTTGGTCAAAAGTT 74  
 |||||  
 RESULT 6  
 BQ866858  
 LOCUS BQ866858 740 bp mRNA linear EST 14-AUG-2002  
 DEFINITION QGC9C05.yg.ab1 QG ABCDI lettuce salinas Lactuca sativa cDNA clone  
 QGC9C05, mRNA sequence.  
 ACCESION BQ866858  
 VERSION BQ866858.1 GI:22252323  
 KEYWORDS EST.  
 SOURCE Lactuca sativa.  
 ORGANISM Lactuca sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asterididae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 Lactuca.  
 1 (bases 1 to 740)  
 REFERENCE Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,  
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison  
 P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,  
 Church,S., Jackson,L. and Bradford,K.  
 Lettuce and Sunflower ESTs from the Compositae Genome Project  
 http://compgenomics.ucdavis.edu/  
 Unpublished (2002)  
 CONTACT: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozikeatgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig QG\_CA\_Contig5488, see http://cgpdb.ucdavis.edu/  
 for details.  
 Plate: QGC9 row: C column: 05.  
 Location/Qualifiers  
 1. .740  
 /organism="Lactuca sativa"  
 /cultivar="Salinas"  
 /db\_xref="taxon:4236"  
 /clone="QGC9C05"  
 /clone\_lib="QG ABCDI lettuce salinas"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that

incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>  
 TAG LIB=QG ABCDI lettuce salinas  
 TAG TISSUE=chemical induction  
 TAG\_SEQ=TGTAGCCGGG"

BASE COUNT 228 a 138 c 165 g 209 t  
 ORIGIN

Query Match 62.4%; Score 25.6; DB 14; Length 740;  
 Best Local Similarity 87.5%; Pred. No. 61;  
 Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 AGGTCAAGGTCATCAAACTAGGTCAAGGT 39  
 |||||  
 DB 43 AGGTCAAGGTCATCAAACTGGGTCAAGTT 74

RESULT 7  
 BQ866155 771 bp mRNA linear EST 14-AUG-2002  
 LOCUS QG7D05.Y9.ab1 QG\_ABCDI lettuce salinas Lactuca sativa cDNA clone  
 DEFINITION QG7D05, mRNA sequence.

ACCESSION BQ866155  
 VERSION BQ866155.1 GI:22251620  
 KEYWORDS EST.

SOURCE Lactuca sativa.  
 ORGANISM Lactuca sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.

REFERENCE 1 (bases 1 to 771)

AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavalle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Stabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.

TITLE Lettuce and Sunflower ESTs from the Compositeae Genome Project

JOURNAL <http://compgenomics.ucdavis.edu/>

COMMENT Unpublished (2002)

Contact: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmunsdon Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659

Email: [akozik@atgc.org](mailto:akozik@atgc.org) [michelmore@vegmail.ucdavis.edu]  
 belongs to contig QG\_CA\_Contig5488, see <http://cgpdb.ucdavis.edu/> for details.

Plate: QG7 row: D column: 05.

FEATURES Location/Qualifiers

1..771

/organism="Lactuca sativa"

/cultivar="Salinas"

/db\_xref="taxon:4236"

/clone="QG7D05"

/clone\_lib="QG\_ABCDI lettuce salinas"

/lab\_host="E.Coli"

/note=vector: pBRCN3AFIAB; The library was constructed from 10 different sources of RNA from a single genotype.

Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>

TAG LIB=QG ABCDI lettuce salinas

TAG TISSUE=chemical induction

TAG\_SEQ=TGTAGCCGGG"

BASE COUNT 240 a 143 c 172 g 216 t  
 ORIGIN

Query Match 62.4%; Score 25.6; DB 14; Length 771;  
 Best Local Similarity 87.5%; Pred. No. 62;  
 Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 AGGTCAAGGTCATCAAACTAGGTCAAGGT 39  
 |||||  
 DB 43 AGGTCAAGGTCATCAAACTGGGTCAAGTT 74

RESULT 8  
 FR0043044 442 bp DNA linear GSS 22-OCT-1999  
 LOCUS Fugu rubripes GSS sequence, clone 159G14AA10, genomic survey  
 DEFINITION Fugu rubripes GSS sequence, clone 159G14AA10, genomic survey sequence.

ACCESSION AL130536

VERSION AL130536.1 GI:16112482

KEYWORDS GSS; genome survey sequence.

SOURCE Takifugu rubripes

ORGANISM Takifugu rubripes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 442)

AUTHORS Elgar, G., Clark, M.S., Smith, S., Meek, S., Warner, S., Edwards, Y.J.K., Umanian, Y., Williams, G. and Brenner, S.

TITLE Direct Submission

JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource

Centre, Hinxton, Cambridge, CB10 1SB. UK Email:

biohelp@hmp.mrc.ac.uk

Vector: pBluescript II KS

V type: phagemid

PRIMER: KS

DBSCR:

One pass dye-terminator sequencing of cosmid cloned genomic

sequence.

FEATURES Location/Qualifiers

1..442

/organism="Takifugu rubripes"

/db\_xref="taxon:3103"

/clone="159G14AA10"

/clone\_lib="cosmid 159G14"

183 a 100 c 84 g 70 t 5 others

BASE COUNT

ORIGIN

Query Match 58.5%; Score 24; DB 17; Length 442;

Best Local Similarity 73.2%; Pred. No. 1.8e+02;

Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAGGTCATCAAACTAGGTCAAGGTCA 41  
 |||||

DB 162 CAATGCCAGAACAAAGGTCAACACGCNAGAACAAAGGTCA 202  
 |||||

RESULT 9

BQ999579 709 bp mRNA linear EST 22-AUG-2002

LOCUS QG22H18.Y9.ab1 QG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone

DEFINITION QG22H18, mRNA sequence.

ACCESSION BQ999579

VERSION BQ999579.1 GI:22433975

KEYWORDS EST.

SOURCE Lactuca sativa.

ORGANISM Lactuca sativa.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.

REFERENCE 1 (bases 1 to 709)

AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,



Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison  
P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,  
Church,S., Jackson,L. and Bradford,K.  
Lettuce and Sunflower ESTs from the Compositae Genome Project  
<http://comgenomics.ucdavis.edu/>  
Unpublished (2002)  
Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: [akozik@ucdavis.edu](mailto:akozik@ucdavis.edu) [[michelmoreveg@mail.ucdavis.edu](mailto:michelmoreveg@mail.ucdavis.edu)]  
belongs to contig QG\_CA\_Contig5488, see <http://cgdb.ucdavis.edu/>  
for details.  
Plate: QGG22 row: H column: 18.

# FEATURES

source

Location/Qualifiers  
1. 709  
/organism="Lactuca sativa"  
/cultivar="L. serriola"  
/db\_xref="taxon:4236"  
/clone="QG22H18"  
/clone\_lib="QG EFGHJ lettuce serriola"  
/lab\_host="E.Coli"

/note="Vector: pBRCNASFIAB; The library was constructed  
from 10 different sources of RNA from a single genotype.  
Separate cDNAs were generated using primers that  
incorporated unique 5' and 3' tags to distinguish each  
source of RNA. cDNAs were then pooled, size-fractionated,  
directionally cloned into a custom medium-copy vector and  
transformations made with four size classes to minimize  
size bias. Details of each source of RNA and library  
construction can be obtained at <http://cgdb.ucdavis.edu/>  
TAG SEQ=Not found"

BASE COUNT 233 a 149 c 161 g 166 t  
ORIGIN

Query Match 58.5%; Score 24; DB 14; Length 709;  
Best Local Similarity 84.4%; Pred. No. 2.2e+02;  
Matches 27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 AGGTCAAGGTCAATCAAACTAGGTCAAGGT 39  
Db 509 AGGTCAAGGTCAATCAAACTAGGTCAAGGT 540

# RESULT 10

AI780559

LOCUS 237 bp mRNA linear EST 18-MAY-2001  
DEFINITION EST261438 tomato susceptible, Cornell Lycopersicon esculentum cDNA  
clone cLES12G22, mRNA sequence.

ACCESSION AI780559

VERSION AI780559.1 GI:5278600

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE 1 (bases 1 to 237)

AUTHORS D' Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E.,  
Liang,F., Upcon,J., Rinning,C.M., Craven,M.B., Fujii,C.Y., Bowman  
,C.B., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,  
Giovannoni,J.J. and Martin,G.B.

TITLE Generation of ESTs from Pseudomonas susceptible tomato

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

# FEATURES

source

Location/Qualifiers  
1. 237

/organism="Lycopersicon esculentum"

/cultivar="Ril-13 (Rio Grande x Money Maker)"

/db\_xref="taxon:4081"

/clone="cLES12G22"

/clone\_lib="tomato susceptible, Cornell"

/tissue\_type="leaf"

/dev\_stage="4-week old"

/lab\_host="SOLR"

/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:

XhoI; cLES - Tomato Pseudomonas Susceptible EST Library.

Directionally cloned cDNAs inserted into pBluescript SK(-)

) at 5' end with EcoRI and 3' end with XhoI site"

BASE COUNT 85 a 36 c 31 g 85 t

# ORIGIN

Query Match 58.0%; Score 23.8; DB 9; Length 237;

Best Local Similarity 80.0%; Pred. No. 1.5e+02;

Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CAAAAGTCAAGGTCAATCAAACTAGGTCAA 35

Db 115 CAAAAGTCAAGGTCAATCAAACTAGGTCAA 149

# RESULT 11

BQ505145/c

LOCUS BQ505145/399 bp mRNA linear EST 22-JUL-2002

DEFINITION EST612560 Generation of a set of potato cDNA clones for microarray  
analyses mixed potato tissues Solanum tuberosum cDNA clone STMGC21  
3' end, mRNA sequence.

ACCESSION BQ505145

VERSION BQ505145.2 GI:21921102

KEYWORDS EST.

SOURCE potato.

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 399)

AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,  
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and  
Karamycheva,S.A.

TITLE Generation of a set of potato cDNA clones for microarray analyses

JOURNAL Unpublished (2002)

COMMENT On Jun 10, 2002 this sequence version replaced gi:21364014.

Contact: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: [potato@tigr.org](mailto:potato@tigr.org)

This clone is available through the Research Genetics, contact the

Research Genetics for further information 1-800-711-6195 or

[cdna@resgen.com](mailto:cdna@resgen.com)

Seq primer: T7.

# FEATURES

source

Location/Qualifiers  
1. 399

/organism="Solanum tuberosum"

/cultivar="Kennebec or Binjite"

/db\_xref="taxon:4113"

/clone="STMGC21"

/clone\_lib="Generation of a set of potato cDNA clones for

microarray analyses mixed potato tissues"

/tissue\_type="mixed tissues"

/lab\_host="SOLR"

/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:

XhoI; supplier: Combination of untreated and Phytophthora

infestans-treated libraries of stolons, leaves, leaflets,

axillary buds of stem explants, petioles, germinating eyes

127 a 83 c 60 g 129 t

# ORIGIN

```

Query Match      58.0%; Score 23.8; DB 14; Length 399;
Best Local Similarity 80.0%; Pred. No. 2e+02;
Matches 28; Conservative 0; Mismatches 0; Indels 7; Gaps 0;

QY 1 CAAACTAGGTCAAAGGTCATCAAAACTAGTCAA 35
    |||||
DB 86 CAAAATACTTCAAAGGCAATCAAAATAGTCAA 52
    |||||

RESULT 12
AW189153/c
LOCUS      AW189153      496 bp      mRNA      linear      EST 22-NOV-1999
DEFINITION x107a10.x1 NCI CGAP Ut4 Homo sapiens cDNA clone IMAGE:2674997 3'
            similar to contains MER27.b1 MER27 MER27 repetitive element ;, mRNA
            sequence.
VERSION     AW189153
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 496)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 418.
            Location/Qualifiers
                source
                1..496
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2674997"
                /clone_lib="NCI CGAP Ut4"
                /tissue_type="serous papillary carcinoma, high grade, 2
                pooled tumors"
                /lab_host="DH10B"
                /note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;
                Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                Average insert size 1.48 kb. Life Technologies catalog #:
                11542-016"

BASE COUNT    136 a 105 c 60 g 195 t
ORIGIN
1..496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2674997"
/clone_lib="NCI CGAP Ut4"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"

Query Match      58.0%; Score 23.8; DB 10; Length 496;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AAACTAGGTCAAAGGTCATCAAAACTAGTCAA 36
    |||||
DB 470 AAAATGGGTAAAGGTGATCAAAACTGTGACAA 436
    |||||

RESULT 13
AW189470/c
LOCUS      AW189470      500 bp      mRNA      linear      EST 22-NOV-1999
DEFINITION x107a10.x1 NCI CGAP Ut4 Homo sapiens cDNA clone IMAGE:2675514 3'
            similar to contains MER27.b1 MER27 MER27 repetitive element ;, mRNA
            sequence.
VERSION     AW189470
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 500)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 415.
            Location/Qualifiers
                source
                1..500
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2675514"
                /clone_lib="NCI CGAP Ut4"
                /tissue_type="serous papillary carcinoma, high grade, 2
                pooled tumors"
                /lab_host="DH10B"
                /note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;
                Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                Average insert size 1.48 kb. Life Technologies catalog #:
                11542-016"

BASE COUNT    140 a 103 c 62 g 195 t
ORIGIN
1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2675514"
/clone_lib="NCI CGAP Ut4"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"

Query Match      58.0%; Score 23.8; DB 10; Length 500;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AAACTAGGTCAAAGGTCATCAAAACTAGTCAA 36
    |||||
DB 473 AAAATGGGTAAAGGTGATCAAAACTGTGACAA 439
    |||||

RESULT 14
BQ868068      373 bp      mRNA      linear      EST 14-AUG-2002
LOCUS      BQ868068
DEFINITION QGD12P06.yg.abi QG ABCDI lettuce salinas Lactuca sativa cDNA clone
            QGD12P06, mRNA sequence.
ACCESSION   BQ868068
VERSION     BQ868068.1 GI:22253825
KEYWORDS    EST.
SOURCE      Lactuca sativa.
ORGANISM    Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
            Lactuca.
            1 (bases 1 to 373)
            Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Riesberg, L.,
            Lin, H., van Damme, M., Lavelle, B., Chevalier, P., Ziegler, J., Ellison
            , P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
            Church, S., Jackson, L. and Bradford, K.
            Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
            Unpublished (2002)
            Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742

```

```

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 500)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 415.
            Location/Qualifiers
                source
                1..500
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2675514"
                /clone_lib="NCI CGAP Ut4"
                /tissue_type="serous papillary carcinoma, high grade, 2
                pooled tumors"
                /lab_host="DH10B"
                /note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;
                Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                Average insert size 1.48 kb. Life Technologies catalog #:
                11542-016"

BASE COUNT    140 a 103 c 62 g 195 t
ORIGIN
1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2675514"
/clone_lib="NCI CGAP Ut4"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"

Query Match      58.0%; Score 23.8; DB 10; Length 500;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AAACTAGGTCAAAGGTCATCAAAACTAGTCAA 36
    |||||
DB 473 AAAATGGGTAAAGGTGATCAAAACTGTGACAA 439
    |||||

RESULT 14
BQ868068      373 bp      mRNA      linear      EST 14-AUG-2002
LOCUS      BQ868068
DEFINITION QGD12P06.yg.abi QG ABCDI lettuce salinas Lactuca sativa cDNA clone
            QGD12P06, mRNA sequence.
ACCESSION   BQ868068
VERSION     BQ868068.1 GI:22253825
KEYWORDS    EST.
SOURCE      Lactuca sativa.
ORGANISM    Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
            Lactuca.
            1 (bases 1 to 373)
            Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Riesberg, L.,
            Lin, H., van Damme, M., Lavelle, B., Chevalier, P., Ziegler, J., Ellison
            , P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
            Church, S., Jackson, L. and Bradford, K.
            Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
            Unpublished (2002)
            Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742

```

Fax: 1-(530)-752-9659  
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig QG\_CA\_Contig6637, see <http://cgpdb.ucdavis.edu/>  
 for details.  
 Plate: QGD12 row: P column: 06.  
 Location/Qualifiers  
 1..373  
 /organism="Lactuca sativa"  
 /cultivar="Salinas"  
 /db\_xref="taxon:4236"  
 /clone="QGD12P06"  
 /clone\_lib="QG ABCDI lettuce salinas"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>  
 TAG\_LIB=QG ABCDI lettuce salinas  
 TAG\_TISSUE=chemical induction  
 TAG\_SEQ=GTGAGCCGGG"

BASE COUNT 101 a 72 c 93 g 107 t  
 ORIGIN  
 Query Match 57.6%; Score 23.6; DB 14; Length 373;  
 Best Local Similarity 86.7%; Pred. No. 2.2e+02;  
 Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

10 GTCAAAGGTCATCAAACTAGGTCAAAGGT 39  
 45 GTAAAGGTCATCAAAATTTGGTCAAAGTT 74

RESULT 15  
 Q862863  
 LOCUS  
 DEFINITION BQ862863 554 bp mRNA linear EST 14-AUG-2002  
 QG222F01.YG.ab1 QG ABCDI lettuce salinas Lactuca sativa cDNA clone  
 QG222F01, mRNA sequence.  
 ACCESSION BQ862863  
 VERSION BQ862863.1 GI:22248338  
 KEYWORDS EST.  
 SOURCE Lactuca sativa.  
 ORGANISM Lactuca sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 Lactuca.  
 1 (bases 1 to 554)  
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
 Lin, H., van Damme, M., Lavelle, D., Chevallier, P., Ziegler, J., Ellison,  
 P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
 Church, S., Jackson, L. and Bradford, K.  
 Lettuce and Sunflower ESTs from the Compositae Genome Project  
<http://composita.ucdavis.edu/>  
 Unpublished (2002)  
 Contact: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig QG\_CA\_Contig6637, see <http://cgpdb.ucdavis.edu/>  
 for details.  
 Plate: QGC22 row: F column: 01.  
 Location/Qualifiers  
 1..554  
 /organism="Lactuca sativa"  
 /cultivar="Salinas"

FEATURES  
 source

/db\_xref="taxon:4236"  
 /clone="QG222F01"  
 /clone\_lib="QG ABCDI lettuce salinas"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>  
 TAG\_LIB=QG ABCDI lettuce salinas  
 TAG\_TISSUE=chemical induction  
 TAG\_SEQ=GTGAGCCGGG"

BASE COUNT 145 a 104 c 138 g 167 t  
 ORIGIN  
 Query Match 57.6%; Score 23.6; DB 14; Length 554;  
 Best Local Similarity 86.7%; Pred. No. 2.7e+02;  
 Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 10 GTCAAAGGTCATCAAACTAGGTCAAAGGT 39  
 DB 48 GTAAAGGTCATCAAAATTTGGTCAAAGTT 77

Search completed: February 20, 2003, 01:08:51  
 Job time : 202.495 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:28:33 ; Search time 161.727 Seconds  
(without alignments)  
9357.426 Million cell updates/sec

Title: US-09-808-388-4

Perfect score: 52

Sequence: 1 caaaactagggtcaaaagtca.....caaaactagggtcaaaagtca 52

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.hcg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.tam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	52	100.0	52	6	AX251576	Sequence
2	28.8	55.4	100168	2	AC108974	Rattus no
3	27	51.9	136120	8	AC068923	Oryza sat
C 4	27	51.9	148002	8	AC079888	Oryza sat
C 5	26.4	50.8	67960	2	AC118695	Mus muscu
C 6	26.2	50.4	194548	10	AL592214	Mouse DNA
C 7	26	50.0	7362	14	TBRN1	Beet kingsp
8	26	50.0	169131	2	AC094713	Rattus no
9	25.6	49.2	175588	2	AC118127	Rattus no
10	25.6	49.2	177071	2	AL772228	Mus muscu
11	25.6	49.2	188201	2	AL808144	Mus muscu
C 12	25.4	48.8	172453	2	AC128896	Rattus no
C 13	25.2	48.5	137286	2	AC025286	Homo sapi
14	25.2	48.5	155491	2	AC117282	Rattus no
C 15	25.2	48.5	188458	9	AC092723	Homo sapi
C 16	25.2	48.5	211001	9	AC018695	Homo sapi
17	25	48.1	170535	9	CNS05T80	Human chr
18	25	48.1	172336	9	CNS06C72	Human chr
C 19	25	48.1	176667	2	AC021720	Homo sapi
C 20	25	48.1	212662	2	AC115730	Mus muscu
C 21	25	48.1	218816	2	AC121571	Mus muscu
C 22	24.8	47.7	191472	2	AC113805	Rattus no
C 23	24.6	47.3	28523	3	U70845	Caenorhabdi
C 24	24.6	47.3	92941	9	AC011313	Homo sapi
C 25	24.4	46.9	10199	9	HUMW11	Human DNA,
C 26	24.4	46.9	20217	1	AE000047	Mycoplasma
C 27	24.4	46.9	67912	2	AC083791	Homo sapi
C 28	24.4	46.9	144454	9	AC009780	Homo sapi
C 29	24.4	46.9	166070	2	AC090871	Oryza sat
C 30	24.4	46.9	185032	2	AC095310	Rattus no
31	24.4	46.9	256116	2	AC106127	Rattus no
32	24.2	46.5	168141	2	AC097807	Rattus no
C 33	24.2	46.5	197298	2	AC113741	Rattus no
C 34	24	46.2	901	4	AF305561	Bos tauru
C 35	24	46.2	75276	9	AC008861	Homo sapi
36	24	46.2	112906	2	AC109755	Rattus no
37	24	46.2	152614	2	AC106737	Homo sapi
38	24	46.2	156554	2	AC099257	Rattus no
39	24	46.2	165707	2	AC104794	Homo sapi
40	24	46.2	168609	2	AC016066	Homo sapi
41	24	46.2	177068	9	AC021504	Homo sapi
C 42	24	46.2	178413	2	AP001370	Homo sapi
C 43	24	46.2	192256	2	AC111987	Rattus no
C 44	24	46.2	193657	2	AC106537	Rattus no
45	23.8	45.8	105930	10	AL606909	Mouse DNA

ALIGNMENTS

RESULT 1  
AX251576  
LOCUS AX251576  
DEFINITION Sequence 4 from Patent WO0168845.  
ACCESSION AX251576  
VERSION AX251576.1 GI:15984999  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1 (bases 1 to 52)  
AUTHORS Massaad, C., Berenbaum, F., Olivier, J.L., Salvat, C. and Bereziat, G.  
TITLE Inflammation-inducible hybrid promoters, vectors containing same  
and uses thereof  
JOURNAL Patent: WO 0168845-A 4 20-SEP-2001;

Aventis Pharma S.A. (FR)  
 Location/Qualifiers  
 1. 52  
 /organism="synthetic construct"  
 /db xref="taxon:32630"  
 /note="element PPPE"  
 19 a 11 c 11 g 11 t  
 BASE COUNT  
 ORIGIN

Query Match 100.0%; Score 52; DB 6; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAACTAGGTCAAAGTCATGCTTTAGGCCAACTAGGTCAAAGGTCA 52  
 Db 1 CAAACTAGGTCAAAGTCATGCTTTAGGCCAACTAGGTCAAAGGTCA 52

RESULT 2  
 AC108974  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-258F1, \*\*\* SEQUENCING IN PROGRESS HTG 13-JUL-2002  
 \*\*\* 45 unordered pieces.  
 AC108974  
 ACCESSION  
 VERSION AC108974.3 GI:21737532  
 KEYWORDS HTG; HTGS PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 100168)  
 Murry D.M., Adams C., Adio-Oduola B., Ali-oman, P.R., Allen C., Albrooks, S.L., Amarating, H.C., Are, J.R., Ayele, M., Banks, T., Bankia, J., Benton, J., Bimge, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, P.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mathiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwunonu, G., Orquyue, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Fu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Roife, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Taney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Unpublished  
 2 (bases 1 to 100168)  
 Worley, K.C.  
 Direct Submission  
 Submitted (03-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 100168)  
 Worley, K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced gi:19846543.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GPHS  
 Center clone name: CH230-258F1  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 55588 bases at least Q40  
 Consensus quality: 58566 bases at least Q30  
 Consensus quality: 61430 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 45 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1001: contig of 1001 bp in length  
 \* 1002 1101: gap of unknown length  
 \* 1102 2247: contig of 1146 bp in length  
 \* 2248 3358: gap of unknown length  
 \* 3359 5150: contig of 1011 bp in length  
 \* 5151 5250: gap of unknown length  
 \* 5251 6256: contig of 1692 bp in length  
 \* 6257 7608: gap of unknown length  
 \* 7609 8967: contig of 1006 bp in length  
 \* 8968 10021: contig of 1159 bp in length  
 \* 10022 10161: gap of unknown length  
 \* 10162 11407: contig of 1094 bp in length  
 \* 11408 12637: contig of 1146 bp in length  
 \* 12638 13803: contig of 1230 bp in length  
 \* 13804 15658: gap of unknown length  
 \* 15659 17175: contig of 1066 bp in length  
 \* 17176 18992: contig of 1755 bp in length  
 \* 18993 20266: gap of unknown length  
 \* 20267 22088: contig of 1417 bp in length  
 \* 22089 23386: contig of 1617 bp in length  
 \* 23387 25266: contig of 1274 bp in length  
 \* 25267 27088: gap of unknown length  
 \* 27089 28187: contig of 1721 bp in length  
 \* 28188 30000: gap of unknown length  
 \* 30001 31199: contig of 1199 bp in length  
 \* 31199 32188: contig of 1199 bp in length

JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

\* 23387 23486: gap of unknown length  
 \* 23487 24763: contig of 1277 bp in length  
 \* 24764 24863: gap of unknown length  
 \* 24864 26063: contig of 1200 bp in length  
 \* 26064 26163: gap of unknown length  
 \* 26164 27457: contig of 1294 bp in length  
 \* 27458 27557: gap of unknown length  
 \* 27558 28799: contig of 1242 bp in length  
 \* 28800 28899: gap of unknown length  
 \* 28900 30446: contig of 1547 bp in length  
 \* 30447 30546: gap of unknown length  
 \* 30547 31109: contig of 2563 bp in length  
 \* 31110 33209: gap of unknown length  
 \* 33210 35469: contig of 2260 bp in length  
 \* 35470 35569: gap of unknown length  
 \* 35570 37801: contig of 2232 bp in length  
 \* 37802 37901: gap of unknown length  
 \* 37902 39834: contig of 1933 bp in length  
 \* 39835 39934: gap of unknown length  
 \* 39935 42002: contig of 2068 bp in length  
 \* 42003 42102: gap of unknown length  
 \* 42103 44937: contig of 2835 bp in length  
 \* 44938 45037: gap of unknown length  
 \* 45038 46779: contig of 1742 bp in length  
 \* 46780 46879: gap of unknown length  
 \* 46880 49607: contig of 2728 bp in length  
 \* 49608 49707: gap of unknown length  
 \* 49709 51317: contig of 1610 bp in length  
 \* 51318 51417: gap of unknown length  
 \* 51418 54254: contig of 2837 bp in length  
 \* 54255 54354: gap of unknown length  
 \* 54355 55768: contig of 1414 bp in length  
 \* 55769 55868: gap of unknown length  
 \* 55869 58459: contig of 2591 bp in length  
 \* 58460 58559: gap of unknown length  
 \* 58560 60503: contig of 2044 bp in length  
 \* 60504 60703: gap of unknown length  
 \* 60704 63401: contig of 2698 bp in length  
 \* 63402 63501: gap of unknown length  
 \* 63502 66194: contig of 2693 bp in length  
 \* 66195 66294: gap of unknown length  
 \* 66295 68437: contig of 2143 bp in length  
 \* 68438 68537: gap of unknown length  
 \* 68538 71580: contig of 3043 bp in length  
 \* 71581 71680: gap of unknown length  
 \* 71681 75628: contig of 3948 bp in length  
 \* 75629 75728: gap of unknown length  
 \* 75729 79844: contig of 4116 bp in length  
 \* 79845 79944: gap of unknown length  
 \* 79945 84166: contig of 4222 bp in length  
 \* 84167 84266: gap of unknown length  
 \* 84267 89708: contig of 5442 bp in length  
 \* 89709 89808: gap of unknown length  
 \* 89809 93101: contig of 3293 bp in length  
 \* 93102 93201: gap of unknown length  
 \* 93202 100168: contig of 6967 bp in length.

## FEATURES

Location/Qualifiers

source

1..100168  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-259F1"

BASE COUNT 27021 a 20887 c 20534 g 27265 t 4461 others  
 ORIGIN

Query Match 55.4%; Score 28.8; DB 2; Length 100169;  
 Best Local Similarity 82.5%; Pred. No. 1.1;  
 Matches 33; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 13 AAAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGTCA 52  
 |||||

Db 85943 AGAGACATGCTTTACCAAACTAGGCAAGGTCA 85982  
 |||||

## RESULT 3

AC068923

LOCUS

DEFINITION

Oryza sativa chromosome 10 BAC OSJNBa0017E08 genomic sequence,

complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

COMMENT

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

ADDRESS

136120 bp DNA linear PLN 17-AUG-2002

Oryza sativa chromosome 10 BAC OSJNBa0017E08 genomic sequence,

complete sequence.

AC068923

AC068923

AC068923.12 GI:22262487

HTG.

Oryza sativa (japonica cultivar-group).

Oryza sativa (japonica cultivar-group).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzoae; Oryza.

1 (bases 1 to 136120)

Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Moffat, K.S., Hall, J.N.,

Gansberger, K., Brenner, M., Burgess, S., Hance, M., Shvartsbeyn, M.,

Tstrin, T., Riggs, F., Hsiao, J., Zismann, V., Blunt, S., Pai, G.,

VanAken, S.E., Utterback, T.R., Feldblyum, T.V., Kalb, E.,

Quackenbush, J., Salzberg, S.L., White, O., and Fraser, C.M.

Oryza sativa chromosome 10 BAC OSJNBa0017E08 genomic sequence

Unpublished

2 (bases 1 to 136120)

Buell, R.

Direct Submission

Submitted (12-MAY-2000) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

3 (bases 1 to 136120)

Buell, R.

Direct Submission

Submitted (04-DEC-2001) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

4 (bases 1 to 136120)

Buell, R.

Direct Submission

Submitted (05-JAN-2002) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org

5 (bases 1 to 136120)

Buell, R.

Direct Submission

Submitted (15-AUG-2002) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org

6 (bases 1 to 136120)

Buell, R.

Direct Submission

Submitted (17-AUG-2002) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org

On Aug 15, 2002 this sequence version replaced gi:17298629.

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org

Address all correspondence to:rice@tigr.org



## KEYWORDS

ORYZA sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 148002)

## AUTHORS

Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Gansberger, K., Kim, M.M.,  
 Overton, II, L.L., Bera, J.J., Tsirnin, T., Krol, M.I., Jarrani, B.B.,  
 Jin, S.S., Koo, H., Ziemann, V., Hsiao, J., Blunt, S., Vanaken, S.S.,  
 Uterback, T.T., Feldblyum, T.V., Yang, Q.Q., Haas, B.J., Suh, B.B.,  
 Peterson, J.J., Quackenbush, J., White, O., Salzberg, S.L. and  
 Fraser, C.M.

## TITLE

Oryza sativa chromosome 10 BAC OSJNBa0078001 genomic sequence

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 148002)

## AUTHORS

Buell, R.

## TITLE

Direct Submission

## JOURNAL

Submitted (16-SEP-2000) The Institute for Genomic Research, 9712

## REFERENCE

Medical Center Dr, Rockville, MD 20850, USA

## AUTHORS

Buell, R.

## TITLE

Direct Submission

## JOURNAL

Submitted (09-AUG-2002) The Institute for Genomic Research, 9712

## REFERENCE

Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org

## COMMENT

On Aug 9, 2002 this sequence version replaced gi:16973742.

Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0078001 is from Oryza sativa chromosome 10

The orientation of the sequence is from SP6 to T7 end of the BAC

clone.

Genes were identified by a combination of several methods: Gene

prediction programs including Fgenesh (<http://www.softberry.com/>),

genscan and Genscan+ (Chris Burge,

<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky,

<http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer

(Mihaila Perte and Steven Salzberg, contact [peptide@tigr.org](mailto:peptide@tigr.org)),

searches of the complete sequence against a peptide database and

the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>).

Annotated genes are named to indicate the level of evidence for

their annotation. Genes with similarity to other proteins are named

after the database hits. Genes without significant peptide

similarity but with EST similarity are named as unknown proteins.

Genes without protein or EST similarity, that are predicted by more

than two gene prediction programs over most of their length are

annotated as hypothetical proteins. Genes encoding tRNAs are

predicted by tRNAscan-SE (Sean Eddy,

<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are

identified by repeatmasker (Arian Smit,

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

This BAC overlaps with rice BACs OSJNBa0017208 (GB:AC068923) and

OSJNBa0073N24 (GB:AC078840).

Location/Qualifiers

1. 148002

/organism="Oryza sativa (japonica cultivar-group)"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="10"

/map="C1361"

/clones="OSJNBa0078001"

/note="japonica cultivar-group"

complement(1..1248)

/gene="OSJNBa0078001.1"

/note="similar to splicing factor PRP6 protein,"

GB:CA486398 GI:559951 (Saccharomyces cerevisiae)"

complement(<1..1248)

/gene="OSJNBa0078001.1"

complement(<1..1248)

/gene="OSJNBa0078001.1"

/codon\_start=1

/product="putative splicing factor, 3'-partial"

/protein\_id="AA093674.1"

/db\_xref="GI:22165057"

## FEATURES

## source

## gene

## mRNA

## CDS

/translation="WVVRAPDGRTHVLDLPSTGLADLTASASRVCGGVPPEQLRL  
 YLAHRLPAPSPSLSLRYSASSLLHLPLGGMTGPTTTPAAPPPPPPSAQP  
 APARYDELNSKPPNIVYAGLRGATGTRSDIGPARAAPDLDRSAAAAAPVGR  
 GKGPFDDGDDGGDEKGYDNQKDFDEGNDAGLFNSADJDDSDRADAWESEI  
 DQMSRRDKREARLKEIEKYSANPKITEQFADLKRLVLDJDAQWESAPIDG  
 SLRNKKKPFESFVPVPTDLEKARQEQSHVTDLPKRAAGGTETPWAQTPVDTLAV  
 GGRGTSLKLDRLSDSVGLTVVDVDPKGLTDLKSMKITSDBAISIDIKKARLLKSV  
 TQTNPKHPGWTAAARLEEVAGKL"  
 complement(716..747)  
 /rpt\_family="(CGA)n"  
 774..803  
 /rpt\_family="(CGG)n"  
 944..983  
 /rpt\_family="(CGG)n"  
 complement(2869..2906)  
 /rpt\_family="(CGA)n"  
 complement(3053..3175)  
 /rpt\_family="(AGGGGG)n"  
 3116..12042  
 /gene="OSJNBa0078001.2"  
 /note="similar to glycine rich protein Tfms GB:CAA64559  
 GI:1166450 (Lycopersicon esculentum); EST BE228432 from  
 this gene"  
 join(3116..3540,4227..4295,4478..4566,4737..4849,  
 5034..5135,5587..5674,5806..6372,6777..7395,7477..7559,  
 8185..8267,8807..9043,9492..9669,9825..10339,10751..10927,  
 11396..11779,11896..12042)  
 /gene="OSJNBa0078001.2"  
 join(3116..3540,4227..4295,4478..4566,4737..4849,  
 5034..5135,5587..5674,5806..6372,6777..7395,7477..7559,  
 8185..8267,8807..9043,9492..9669,9825..10339,10751..10927,  
 11396..11779,11896..12042)  
 /gene="OSJNBa0078001.2"  
 /codon\_start=1  
 /product="putative glycine-rich protein"  
 /protein\_id="AA093675.1"  
 /db\_xref="GI:22165058"  
 /translation="WTSSAPNPSPIVPTPTPTANAAAAAGAVPVSPPTPPKD  
 QQEGGGGGVGDGVEEVEVGGEEMEDVGGAGGGGGVGGVGGGGGAGGG  
 GGGGGQQASPAFVIRILKQPPSRHKRVFELCRNFSVAVWCKKLNACASETC  
 ARIPTSSPPFWPIHILNPERPTCSFVNKADSPDFVQFISPRSCPRALLVA  
 NFGRITITWTQGTPTNLVDASSQCEHEWRQDLVTVTKLGSIGISYRWLPANSST  
 SNNLITFEKFLTQQPSFSSGSLQHSQWSPQSAQPRFSTKGLPAGPSGI  
 MAADLTETGALHGVPLNPSVTVMVMPGLNGIQATAKINATSSLPNPSI  
 LHWAPLASYLFSLODYLVSQAQTKQAQVDNETTVASHTCCPVSFSAVYSP  
 SSSECSVDNRLSFDPLDNDVRLQAQIVYSAHGEVAVAFRGVHIFSGNFEQ  
 VDSYHNVGSAIAPPAFSSGCLASVWHDTLKORTILKIIVLPAILNACTKVSSA  
 WBERADSRFWMLLAGVDMWDVAGCTQSAEDGIVSLNSVIALDLDFCLTIOQ  
 ROOHCNLDRIKCRLLGNTAODRALVDMQARLLDMLGKIEBALINFTLLPEP  
 WQASSMLSSIGPDWTVDPALLSTIQGVDAVLDSLASHFTLRRLRYAFCTLASHT  
 VGASSGPNRNMVTSPTNSPSTNQGQGVASTTGSOMQEWOGATAKLSNNT  
 DGAANAAPISGRSFMPISTINTGTFPTPAVRLIGDCHFLHLCQLLCLLFRRR  
 QSPRIANAQKSDSMQHLNMTEDNTLAVRSLGAKLEDTGTSRQMGVAGK  
 AENPVNKSARIGSNAGQYTSDEVKVLFLVLDCKRTATLPHLPFSQVGSNN  
 IIRLHYDGNVTVLPVWEASLGPNQNPMPRGDAAGLLRELEQAPAEWHRRN  
 MFGQWSEDDLPDNTQLKINGSTNRHLSDEEDGDSFGIQLNWRKRISERD  
 AAFGLTSVGLSGFLGWSRSDVATVWTKTGLEWYKICIRLAQTCAFAQPGALAP  
 NTSNELEAWISRWTHACXCCGTVKVVV"  
 complement(3177..3240)  
 /rpt\_family="(CGG)n"  
 3266..3439  
 /rpt\_family="(TGGGG)n"  
 complement(5302..5332)  
 /rpt\_family="AT\_rich"  
 complement(13017..13071)  
 /rpt\_family="AT\_rich"  
 13524..13562  
 /rpt\_family="(GGGA)n"  
 15031..15173  
 /rpt\_family="tandem"  
 15224..20358

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## gene





*	9740	9839:	gap of	100	bp
*	9840	10568:	contig of 729	bp	in length
*	10569	10668:	gap of	100	bp
*	10669	11390:	contig of 722	bp	in length
*	11391	11490:	gap of	100	bp
*	11491	12185:	contig of 635	bp	in length
*	12186	12285:	gap of	100	bp
*	12286	13002:	contig of 717	bp	in length
*	13003	13102:	gap of	100	bp
*	13103	13776:	contig of 674	bp	in length
*	13777	13876:	gap of	100	bp
*	13877	14587:	contig of 711	bp	in length
*	14588	14689:	gap of	100	bp
*	14688	15418:	contig of 731	bp	in length
*	15419	15516:	gap of	100	bp
*	15519	16234:	contig of 716	bp	in length
*	16235	16334:	gap of	100	bp
*	16335	17053:	contig of 719	bp	in length
*	17054	17153:	gap of	100	bp
*	17154	17890:	contig of 737	bp	in length
*	17891	17990:	gap of	100	bp
*	17991	18729:	contig of 739	bp	in length
*	18730	18829:	gap of	100	bp
*	18830	19559:	contig of 730	bp	in length
*	19560	19659:	gap of	100	bp
*	19660	20396:	contig of 737	bp	in length
*	20397	20496:	gap of	100	bp
*	20497	21231:	contig of 735	bp	in length
*	21232	21331:	gap of	100	bp
*	21332	22058:	contig of 727	bp	in length
*	22059	22158:	gap of	100	bp
*	22159	22874:	contig of 716	bp	in length
*	22875	22974:	gap of	100	bp
*	22975	23694:	contig of 720	bp	in length
*	23695	23794:	gap of	100	bp
*	23795	24533:	contig of 739	bp	in length
*	24534	24633:	gap of	100	bp
*	24634	25360:	contig of 727	bp	in length
*	25361	25460:	gap of	100	bp
*	25461	26176:	contig of 716	bp	in length
*	26177	26276:	gap of	100	bp
*	26277	26989:	contig of 713	bp	in length
*	26990	27089:	gap of	100	bp
*	27090	27828:	contig of 739	bp	in length
*	27829	27928:	gap of	100	bp
*	27929	28653:	contig of 725	bp	in length
*	28654	28753:	gap of	100	bp
*	28754	29490:	contig of 737	bp	in length
*	29491	29590:	gap of	100	bp
*	29591	30306:	contig of 716	bp	in length
*	30307	30406:	gap of	100	bp
*	30407	31117:	contig of 711	bp	in length
*	31118	31217:	gap of	100	bp
*	31218	31932:	contig of 715	bp	in length
*	31933	32032:	contig of 715	bp	in length
*	32033	32747:	contig of 715	bp	in length
*	32748	32847:	gap of	100	bp
*	32848	33559:	contig of 712	bp	in length
*	33560	33659:	gap of	100	bp
*	33660	34361:	contig of 702	bp	in length
*	34362	34461:	gap of	100	bp
*	34462	35179:	contig of 718	bp	in length
*	35180	35799:	gap of	100	bp
*	35799	36011:	contig of 732	bp	in length
*	36012	36111:	gap of	100	bp
*	36112	36820:	contig of 709	bp	in length
*	36821	36920:	gap of	100	bp
*	36921	37653:	contig of 733	bp	in length
*	37654	37753:	gap of	100	bp
*	37754	38475:	contig of 722	bp	in length
*	38476	38575:	gap of	100	bp
*	38576	39316:	contig of 741	bp	in length
*	39317	39416:	gap of	100	bp

Query Match	Score	DB 2	Length
Best Local Similarity	69.8%		
Matches	36	Conservative	16
Query	1	CAAAAGCTAGGTC	52
Subject	38129	CAGACCTTACGTTAAGACACATATCTGTAGGCAGCAAAATATTTTCCATAGGCA	380708

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., ENBL; Sw., SWISSPROT; Ir., TREMBL; Wp., WORMPEP; information on the WORMPEP database can be found at

[http://www.ranger.ac.uk/projects/C\\_elegans/wormpep](http://www.ranger.ac.uk/projects/C_elegans/wormpep) RP23-212124 is from the RP01-23 Mouse PAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6. Location/Qualifiers

FEATURES  
source

1. 194548

/organism="Mus musculus"

/db\_xref="taxon:10090"

/chromosome="1"

/clone="RP23-212124"

/clone\_lib="RP01-23"

BASE COUNT 61316 a 35910 c 35045 g 62277 t

ORIGIN

Query Match 50.4%; Score 26.2; DB 10; Length 194548;

Best Local Similarity 72.3%; Pred. No. 11;

Matches 34; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy 1 CAAACTAGGTCAGAGTCATGCTTTAGGCCCAAACTAGGTCAAA 47

Db 14017 CAATACCAAGCATAGTGTGTCCTAGGACTAGACTGGCTCAA 13971

RESULT 7

TERRNA1/c

LOCUS TERRNA1 Beet ringspot virus gene for polyprotein, complete cds. 7362 bp RNA linear VRL 29-MAY-2002

DEFINITION Beet ringspot virus gene for polyprotein, complete cds.

ACCESSION D00322

VERSION D00322.1 GI:538250

KEYWORDS RNA-1; TBRV.

SOURCE Beet ringspot virus (strain:S) cDNA to genomic RNA.

ORGANISM Beet ringspot virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae; Nepovirus; Subgroup B.

REFERENCE

AUTHORS

TITLE Greif,C., Hemmer,O. and Fritsch,C.

JOURNAL Nucleotide sequence of tomato black ring virus RNA-1

REFERENCE J. Gen. Virol. 69, 1517-1529 (1988)

AUTHORS 2 (bases 1 to 7362)

JOURNAL Fritsch,C.

COMMENT Unpublished

On Sep 15, 1994 this sequence version replaced gi:222642.

TBRV has a genome consisting of two ssRNA species of RNA-1 and

RNA-2. The sequence of the RNA-1 was determined from overlapping

cDNA clones

FEATURES

source

1. 7362

/organism="Beet ringspot virus"

/strain="S"

/db\_xref="taxon:191547"

/chromosome="RNA-1"

/note="5' end of TBRV RNA-1--clones p1831, p1628, p103,

p1792, p172, p160, psrF29"

261..7061

/note="ORF"

/codon\_start=1

/product="polyprotein"

CDS

/protein\_id="BAA00234.1"  
/db\_xref="GI:538251"  
/translation="MSVTLSPSGDGFSPNHVKNNSLNKLYFYNSNLDIVLDDDFDYF  
NRYVKNVLLSFFSDRVLSALYTSMSVSEASLAMEDFCFELALDELKINPFHOLWEE  
TIANWVPVPGSLDFCTOYEIRREARASAEILRLKEVARORAFDEVEFLKHGA  
KHPAPSAQAQRAGKQKKGILLGKNNKAKALGAHSAVARAQAKAEVLREPE  
PSPQIQRALEAQIFADRLSKYALTAARVAKAAARELREKELFLVTDQLLNAPL  
PMEVQKERKYRVPRTGTSNTTPKPNVLELCPFMGLGAKTADRETQILNAGSI  
HQYRSLASAIYAWVLGSPMFCIAPVKTIFKLTFTWVDFPEEVLIDELNKNSA  
RCFEASLVLEERAKLEAHFANCRANIFMKAMAGVQMAKCAYSGLFTCEEAGRS  
LSEGVSWINSFRECIMHKELGACMLTVMKIKVKNVNSMLKHLGCLATLGT  
YAWVALLLGLLTLLERCIGGAGILTLFVTGVEFAAICGLHAAGSEGLQREVMQ  
CTALAGPDPVHKNGKSPWADILAEORLEDRADNRVRSIPISGLISMQQPGG  
LCSMSVLEIGKUGAACHSRMGKEALKEFCATQMLTIGRISDKVTRGVTFDEL  
STLVSDVRGWLCAACRESFTEIGNQFCDWVAQLVDDQKQVGVNGVPRKLS  
TYSGLIQIMKDMELHKTIRAGISSEGRCEPWTYLFQRCGSGNSFNATLDNAL  
KHFGQIPNTAYRCKDSFFGYSGOTFFRVDLSSVKLDPMBEAMINLVSCOEVL  
NADLADPIYERSPFISSNFEDVPAGCGVDIEAYRSKACLVEMRRKPGVLPD  
DNPLASORAKDPMSONLLESGOTENSMLMEDVTEIINISARHRAAOKLQARYM  
REKSLDLPLAAENFLGEVQNAVLDPSGLIELEKAGIPKQGGSGLYVDSGLYLLP  
TFQFEIPIITDDGYKRLWDERNRKSFSLKIQIOTGTYLNSKNVVTGFLRSLVNGCAVL  
SKDTLSSASAAQGSIFKALGIDRIYLRTOHQDLDSADIPENPYSNSAWIKLKA  
IGMERTYLAENGCGIIMIAAALILVSAWGFKLFTGLFSGMSLGAIVGSADV  
KQKQSSQGGYRARNIPHRYAYAKSQAGDGLPAARLCVAIYQGGGFVSAMQY  
KNKSVTRHOLRFOEGEOLTVIPSTSGESQLRWHKYHMRPEPGEIVTWLAPSLP  
SLSPDLKDLFLKEDKEVDLPNHEKTIGYLRVDNTAFHVDLDTYAADKTPLPKGVV  
GNELYHSITPEKITHYESRNDGCMILCOIKGMBVNGMLVAGDKDKTSWADIMPEN  
THAELOQIEYIPKFGAYDGFPGAGYVPMADFTPLPKNNVNPVQSLRVPDCDPIK  
EPAVLNADKRCFAGNPFVTKKKTFHFMKELEQEIIDEVATDIILETYDCEDHYL  
NDIPLVAINGPADSEAELENFVMTSPGYFYKNNRAEKLKGSAYPEEAEEDGYL  
KLKRGMAAKHENLVETPNEVPELVETCTDELLPERIKVGCALFEIMPLHYN  
LFLRQKTCATQFLOHNRHVLPCOVTGTPYSRWGHMLNLRMPKTNKNEALCDSYSGFD  
GLNQAQVTEIAKM.NRLYALSGSEVQAOYNNLMALVGYAFVQPEVYKVCGLP  
SGFALTYVNSVNEILLIRYAKLAPKPERNRNQVCLLVYGDNLISVSPSIASW  
FTEGIRITLASKKVKITDGDOKRPTTEAKSPWELDFKRLKLLKDNGLVQAEIDRS  
AIFSSLYMLTPDKSFHASQASQDQGTVDVVEELINNVNVALMELYLNDPREFSRV  
RDYIKALPLATQPTWAFCEAFHSAQQTGMKLYDKPAKLDHMSGLDFKFKPMHVSQV  
GNKAHYFTMLGVAGPHYKPOENDPITVSTPELKMVGCEHVPIQYSGVGLGPTKKWY  
LDGRPSOLKNKLYIHPILRAQIEAGKRLVFNSPAPYVANNALAFGTGGKMLIQ  
KDALVHRYNVAPESTSGLEQYEDAPLPTATIGTFVFANGETYAALCEYKEGKVLNYEG  
PFTILINENAKDRKYPQWATQAKTKFKVSLACDSTWCPHTVCEYKAFRCWLA  
KOKTSARKVSPWHGTKLS"

894

/citation=[1]

/replace="c"

2498

/citation=[1]

/replace=""

2502

/citation=[1]

/replace=""

2510..2511

/citation=[1]

/replace="ac"

2525

/citation=[1]

/replace=""

3241

/citation=[1]

/replace=""

3277

/citation=[1]

/replace=""

3293

/citation=[1]

/replace=""

3992

/citation=[1]

/replace="a"

4014

/citation=[1]

/replace=""

4112

/replace=""

4112

/replace=""

4112

/replace=""

4112

/replace=""

4112

/replace=""

4112

```

/citation=[1]
/replace="c"
7362
/polya_site
BASE COUNT 1927 a 1537 c 1734 g 2164 t
ORIGIN
Query Match 50.0%; Score 26; DB 14; Length 7362;
Best Local Similarity 76.2%; Pred. No. 16;
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
2Y 2 AAAGTAGGTCAGAGGTCATGCTTTAGGCCAACAACTAGGT 43
2b 5304 AGAAGTGGGTGAAGACATGCTTTTGGCGCAAAAGAGAT 5263

RESULT 8
AC094713 169131 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-5M3, *** SEQUENCING IN PROGRESS ***
ACCESSION AC094713
VERSION AC094713.2 GI:17941492
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 169131)
REFERENCE
AUTHORS Mueny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amarutunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Briefe,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,D., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Louissegue,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenwo,S.,
Ogun,M., Okwomu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,F., Telford,B., Thomas,N.,
Thomas,S., Umami,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleciyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 169131)
AUTHORS Worley,K.C.

```

```

TITLE Direct Submission
JOURNAL Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Dec 20, 2001 this sequence version replaced gi:15624549.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBPK
Center clone name: CH230-5M3
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 137339 bases at least Q40
Consensus quality: 145334 bases at least Q30
Consensus quality: 152478 bases at least Q20
Estimated insert size: 144928; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* consists of 72 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 6147: contig of 6147 bp in length
6148 6247: gap of unknown length
6248 11125: contig of 4878 bp in length
11126 11225: gap of unknown length
11226 15620: contig of 4395 bp in length
15621 15720: gap of unknown length
15721 20107: contig of 4387 bp in length
20108 20207: gap of unknown length
20208 23694: contig of 3487 bp in length
23695 23794: gap of unknown length
23795 29170: contig of 5376 bp in length
29171 32514: contig of 3244 bp in length
32515 32614: gap of unknown length
32615 36226: contig of 3612 bp in length
36227 36326: gap of unknown length
36327 40109: contig of 3783 bp in length
40110 40209: gap of unknown length
40210 44265: contig of 4056 bp in length
44266 44366: gap of unknown length
44367 46398: contig of 2033 bp in length
46399 46498: gap of unknown length
46499 50131: contig of 3633 bp in length
50132 50231: gap of unknown length
50232 52689: contig of 2458 bp in length
52690 52789: gap of unknown length
52790 55275: contig of 2486 bp in length
55276 55375: gap of unknown length
55376 59123: contig of 3748 bp in length
59124 59223: gap of unknown length
59224 61657: contig of 2434 bp in length
61658 61758: gap of unknown length
61759 64006: contig of 2248 bp in length
64007 67885: gap of unknown length
67886 69632: contig of 1647 bp in length
69633 69732: gap of unknown length
69733 72663: contig of 2831 bp in length
72664 72564: gap of unknown length

```

TITLE  
JOURNAL  
COMMENT

72664 75289: contig of 2626 bp in length  
 75290 75389: gap of unknown length  
 75390 78077: contig of 2688 bp in length  
 78078 78177: gap of unknown length  
 78178 79775: contig of 1798 bp in length  
 79776 80075: gap of unknown length  
 80076 82495: contig of 2420 bp in length  
 82496 82595: gap of unknown length  
 82596 84301: contig of 1706 bp in length  
 84302 84401: gap of unknown length  
 84402 85860: contig of 1459 bp in length  
 85861 85960: gap of unknown length  
 85961 87989: contig of 2029 bp in length  
 87990 88089: gap of unknown length  
 88090 89005: contig of 1816 bp in length  
 89006 90005: gap of unknown length  
 90006 92507: contig of 2502 bp in length  
 92508 92607: gap of unknown length  
 92608 95234: contig of 2627 bp in length  
 95235 95334: gap of unknown length  
 95335 96788: contig of 2344 bp in length  
 96789 97778: gap of unknown length  
 97779 100227: contig of 2449 bp in length  
 100228 100327: gap of unknown length  
 100328 102039: contig of 1712 bp in length  
 102040 102139: gap of unknown length  
 102140 104873: contig of 2734 bp in length  
 104874 104973: gap of unknown length  
 104974 107341: contig of 2368 bp in length  
 107342 107441: gap of unknown length  
 107442 110213: contig of 2774 bp in length  
 110216 110315: gap of unknown length  
 110316 113182: contig of 2867 bp in length  
 113183 113283: gap of unknown length  
 113284 115053: contig of 1771 bp in length  
 115054 115153: gap of unknown length  
 115154 117117: contig of 1963 bp in length  
 117117 119376: contig of 1160 bp in length  
 119377 118476: gap of unknown length  
 118477 120665: contig of 2189 bp in length  
 120666 120765: gap of unknown length  
 120766 122034: contig of 1269 bp in length  
 122035 122133: gap of unknown length  
 122134 123613: contig of 1479 bp in length  
 123614 123713: gap of unknown length  
 123714 125903: contig of 2190 bp in length  
 125904 126003: gap of unknown length  
 126004 127528: contig of 1525 bp in length  
 127529 127628: gap of unknown length  
 127629 129390: contig of 1782 bp in length  
 129391 129491: contig of unknown length  
 129492 130887: contig of 1397 bp in length  
 130888 130987: gap of unknown length  
 130988 132177: contig of 1090 bp in length  
 132178 133683: contig of 1506 bp in length  
 133684 133783: gap of unknown length  
 133784 135651: contig of 1868 bp in length  
 135652 137153: gap of unknown length  
 137154 137253: contig of 1401 bp in length  
 137254 138382: contig of 1130 bp in length  
 138383 138482: gap of unknown length  
 138483 140513: contig of 2031 bp in length  
 140514 140613: gap of unknown length  
 140614 142035: contig of 1422 bp in length  
 142036 142135: gap of unknown length  
 142136 143802: contig of 1667 bp in length  
 143803 143903: gap of unknown length  
 143904 145113: contig of 1211 bp in length  
 145114 145213: gap of unknown length

Query Match 50.0%; Score 26; DB 2; Length 169131;  
 Best Local Similarity 70.0%; Pred. No. 13;  
 Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAGAGGTCATGCTTTAGGCGCCAAAGCTAGGTCAGAGGT 50  
 Db 118953 CAGGCAAGTGATAGGACATATCTATAGATCAGAAAGCAGGTCAAAGTT 119002

RESULT 9  
 AC118127

LOCUS Rattus norvegicus clone CH230-155F18, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 59 unordered pieces.  
 AC118127 175588 bp DNA linear HTG 23-JUL-2002

ACCESSION AC118127.4 GI:21903182

KEYWORDS HTG; HTGS PHASE1

SOURCE Rattus norvegicus

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 175588)

AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Alebrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, T.,  
 Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowie, S., Brivea, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denny, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
 Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, D.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogun, M., Okwodu, G.,  
 Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, I.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tausey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczky, R., Woodson, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 175588)

Worley, K.C.

Direct Submission

Submitted (13-APR-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 175588)

Worley, K.C.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

TITLE Direct Submission  
 JOURNAL Submitted (23-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 COMMENT On Jul 18, 2002 this sequence version replaced gi:20260763.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GWHQ  
 Center clone name: CH230-155F18  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 117684 bases at least Q40  
 Consensus quality: 123861 bases at least Q30  
 Consensus quality: 128719 bases at least Q20  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 59 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1  
 1797 1796: contig of 1796 bp in length  
 1896: gap of unknown length  
 1897 3162: contig of 1266 bp in length  
 3163 3262: gap of unknown length  
 3263 4399: contig of 1137 bp in length  
 4400 4499: gap of unknown length  
 4500 6066: contig of 1567 bp in length  
 6067 6166: gap of unknown length  
 6167 7309: contig of 1143 bp in length  
 7310 7409: gap of unknown length  
 7410 8999: contig of 1590 bp in length  
 9000 9099: gap of unknown length  
 9100 10614: contig of 1515 bp in length  
 10615 10714: gap of unknown length  
 10715 11787: contig of 1073 bp in length  
 11788 11887: gap of unknown length  
 11888 13613: contig of 1726 bp in length  
 13614 13713: gap of unknown length  
 13714 15160: contig of 1447 bp in length  
 15161 16335: contig of 1075 bp in length  
 16336 16435: gap of unknown length  
 16436 18239: contig of 1804 bp in length  
 18240 18339: gap of unknown length  
 18340 19643: contig of 1304 bp in length  
 19644 19743: gap of unknown length  
 19744 20903: contig of 1160 bp in length  
 20904 21003: gap of unknown length  
 21004 22897: contig of 1894 bp in length  
 22898 22977: gap of unknown length  
 22978 23997: contig of 1200 bp in length  
 23998 25169: gap of unknown length  
 25170 25269: gap of unknown length  
 25270 26969: gap of unknown length  
 26970 28264: contig of 1295 bp in length  
 28265 28364: gap of unknown length  
 28366 29336: contig of 1072 bp in length  
 29337 31095: contig of 1559 bp in length  
 31096 31195: gap of unknown length

\* 31196 32533: contig of 1338 bp in length  
 32534 32633: gap of unknown length  
 32634 35386: contig of 2753 bp in length  
 35387 35486: gap of unknown length  
 35487 37291: contig of 1805 bp in length  
 37292 38662: contig of 1271 bp in length  
 38663 38762: gap of unknown length  
 38763 40312: contig of 1550 bp in length  
 40313 42126: contig of 1714 bp in length  
 42127 42286: gap of unknown length  
 42287 44089: contig of 1863 bp in length  
 44090 44189: gap of unknown length  
 44190 47095: contig of 2906 bp in length  
 47096 47195: gap of unknown length  
 47196 48962: contig of 1767 bp in length  
 48963 49062: gap of unknown length  
 49063 51181: contig of 2119 bp in length  
 51182 51281: gap of unknown length  
 51282 53965: contig of 2684 bp in length  
 53966 56853: contig of 2788 bp in length  
 56854 56953: gap of unknown length  
 56954 59509: contig of 2556 bp in length  
 59510 62483: gap of unknown length  
 62484 62589: gap of unknown length  
 62590 65206: contig of 2637 bp in length  
 65207 65306: gap of unknown length  
 65307 67901: contig of 2595 bp in length  
 67902 68001: gap of unknown length  
 68002 70944: contig of 2943 bp in length  
 70945 71044: gap of unknown length  
 71045 73973: contig of 2929 bp in length  
 73974 74073: gap of unknown length  
 74074 76611: contig of 2538 bp in length  
 76612 76711: gap of unknown length  
 76712 79580: contig of 2869 bp in length  
 79581 79680: gap of unknown length  
 79681 82571: contig of 2891 bp in length  
 82572 82671: gap of unknown length  
 82672 85735: contig of 3064 bp in length  
 85736 85835: gap of unknown length  
 85836 89400: contig of 3565 bp in length  
 89401 93548: contig of 4048 bp in length  
 93549 93648: gap of unknown length  
 93649 97229: contig of 3581 bp in length  
 97230 97329: gap of unknown length  
 97330 101176: contig of 3847 bp in length  
 101177 101276: gap of unknown length  
 101277 105139: contig of 3863 bp in length  
 105140 105239: gap of unknown length  
 105240 109241: contig of 4002 bp in length  
 109242 109341: gap of unknown length  
 109342 113657: contig of 4316 bp in length  
 113658 113757: gap of unknown length  
 113758 117861: contig of 4104 bp in length  
 117862 117961: gap of unknown length  
 117962 123522: contig of 5561 bp in length  
 123523 123622: gap of unknown length  
 123623 130088: contig of 6466 bp in length  
 130089 130188: gap of unknown length  
 130189 135446: contig of 5258 bp in length

Query Match 49.2%; Score 25.6; DB 2; Length 175588;  
 Best Local Similarity 70.8%; Pred. No. 19;  
 Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

CY 5 ACTAGGTCAAGGTCTGTTTATGAGCCCAAACTAGGTCAAGGTCA 52  
 Db 105560 AGTAGGTCAAGGTCTGTTTATGAGTCACAAACCAAGGTCTGTGA 105607

```

RESULT 10
AL772228      177071 bp DNA linear HTG 17-AUG-2002
LOCUS        Mus musculus chromosome 4 clone RP23-236D1, *** SEQUENCING IN
DEFINITION   PROGRESS ***, 4 unordered pieces.
ACCESSION    AL772228
VERSION      AL772228
KEYWORDS     HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE       house mouse.
ORGANISM     Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE        Lovell, J.
JOURNAL      Direct Submission
COMMENT      Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
             On Aug 21, 2002 this sequence version replaced gi:22204574.
             ----- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: humquery@sanger.ac.uk
             ----- Project Information
             Center project name: BM236D1
             ----- Summary Statistics
             Assembly program: XGAP4; version 4.5
             Chemistry: Dye-terminator; 98% of reads
             Chemistry: Dye-primer Big Dye; 0% of reads
             Consensus quality: 176172 bases at least Q40
             Consensus quality: 176430 bases at least Q30
             Consensus quality: 176639 bases at least Q20
             Insert size: 176771; sum-of-contigs
             Insert size: 172123; 2.8% error; agarose-fp
             Quality coverage: 9.16x in Q20 bases; sum-of-contigs Quality
             coverage: 9.94x in Q20 bases; agarose-fp
             -----
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 4 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence.
             * as soon as it is available and the accession number will
             * be preserved.
             * 1 167555: contig of 167555 bp in length
             * 167556 167655: gap of 100 bp
             * 167656 170395: contig of 2740 bp in length
             * 170396 170495: gap of 100 bp
             * 170496 172644: contig of 2149 bp in length
             * 172645 172744: gap of 100 bp
             * 172745 177071: contig of 4327 bp in length.
             Location/Qualifiers
               1..177071
                 /organism="Mus musculus"
                 /db_xref="taxon:10090"
                 /chromosome="4"
                 /clone="RP23-236D1"
                 /clone_lib="RPCI-23"
             misc_feature
               1..167555
                 /note="assembly fragment:00431"
             misc_feature
               167656..170395
                 /note="assembly fragment:02163"
             misc_feature
               170496..172644
                 /note="assembly fragment:02216"
             misc_feature
               172745..177071
                 /note="assembly fragment:02884"
             BASE COUNT 49993 a 36060 c 38312 g 52406 t 300 others
ORIGIN

```

```

Query Match 49.2%; Score 25.6; DB 2; Length 177071;
Best Local Similarity 77.5%; Pred. No. 19;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 13 AAGGTCATCTTTAGGCCCAAACTAGTCAAAAGGTCA 52
      |||||
Db 16453 AGAGCATCTATTCACCAAAACTAGTCAAAAGGTCA 16492
      |||||

RESULT 11
AL808144      188201 bp DNA linear HTG 09-JUL-2002
LOCUS        Mus musculus chromosome 4 clone RP23-157G9, *** SEQUENCING IN
DEFINITION   PROGRESS ***, 24 unordered pieces.
ACCESSION    AL808144
VERSION      AL808144.2 GI:21727636
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE       house mouse.
ORGANISM     Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE        Sims, S.
JOURNAL      Direct Submission
COMMENT      Submitted (08-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
             ----- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: humquery@sanger.ac.uk
             ----- Project Information
             Center project name: BM157G9
             ----- Summary Statistics
             Assembly program: XGAP4; version 4.5
             Chemistry: Dye-terminator; 100% of reads
             Consensus quality: 180988 bases at least Q40
             Consensus quality: 183317 bases at least Q30
             Consensus quality: 184774 bases at least Q20
             Insert size: 185901; sum-of-contigs
             Insert size: 182559; 3.9% error; agarose-fp
             Quality coverage: 5.03x in Q20 bases; sum-of-contigs Quality
             coverage: 5.12x in Q20 bases; agarose-fp
             -----
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 24 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence.
             * as soon as it is available and the accession number will
             * be preserved.
             * 1 9010: contig of 9010 bp in length
             * 9011 9110: gap of 100 bp
             * 9111 16086: contig of 6986 bp in length
             * 16087 16186: gap of 100 bp in length
             * 16197 23899: contig of 9703 bp in length
             * 25900 25999: gap of 100 bp
             * 26000 35140: contig of 9141 bp in length
             * 35141 35240: gap of 100 bp
             * 35241 39016: contig of 3776 bp in length
             * 39017 39116: gap of 100 bp
             * 39117 42913: contig of 3797 bp in length
             * 42914 43013: gap of 100 bp
             * 43014 55275: contig of 12262 bp in length
             * 55276 55375: gap of 100 bp
             * 55376 60048: contig of 4673 bp in length
             * 60049 60148: gap of 100 bp
             * 60149 71395: contig of 11247 bp in length
             * 71396 71495: gap of 100 bp
             * 71496 74860: contig of 3365 bp in length
             * 74861 74960: gap of 100 bp

```

\* 74961 97774: contig of 22814 bp in length  
\* 97775 97874: gap of 100 bp  
\* 97875 103188: contig of 5314 bp in length  
\* 103189 103288: gap of 100 bp  
\* 103289 117057: contig of 13769 bp in length  
\* 117058 117157: gap of 100 bp  
\* 117158 121704: contig of 4547 bp in length  
\* 121705 121804: gap of 100 bp  
\* 121805 123958: contig of 2154 bp in length  
\* 123959 124058: gap of 100 bp  
\* 124059 129614: contig of 5556 bp in length  
\* 129615 129714: gap of 100 bp  
\* 129715 132913: contig of 3199 bp in length  
\* 132914 133013: gap of 100 bp  
\* 133014 141267: contig of 8254 bp in length  
\* 141268 141367: gap of 100 bp  
\* 141368 147593: contig of 6226 bp in length  
\* 147594 147693: gap of 100 bp  
\* 147694 159744: contig of 12051 bp in length  
\* 159745 159844: gap of 100 bp  
\* 159845 170736: contig of 10892 bp in length  
\* 170737 170836: gap of 100 bp  
\* 170837 173576: contig of 2740 bp in length  
\* 173577 173676: gap of 100 bp  
\* 173677 185245: contig of 11569 bp in length  
\* 185246 185345: gap of 100 bp  
\* 185346 188201: contig of 2856 bp in length.  
FEATURES  
Source  
1..188201  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-157G9"  
/clone\_lib="RPCI-23"  
1..3010  
/note="assembly\_fragment:01290"  
fragment\_chain:1  
clone\_end:SP6  
vector\_side:left  
9111..16096  
/note="assembly\_fragment:01608"  
fragment\_chain:1  
16197..25899  
/note="assembly\_fragment:00497"  
fragment\_chain:1  
26000..35140  
/note="assembly\_fragment:01736"  
fragment\_chain:1  
35241..39016  
/note="assembly\_fragment:01225"  
fragment\_chain:1  
39117..42913  
/note="assembly\_fragment:00697"  
fragment\_chain:2  
43014..55275  
/note="assembly\_fragment:01538"  
fragment\_chain:2  
55376..60048  
/note="assembly\_fragment:01446"  
fragment\_chain:2  
60149..71395  
/note="assembly\_fragment:00475"  
fragment\_chain:2  
71496..74860  
/note="assembly\_fragment:00149"  
fragment\_chain:2  
74961..97774  
/note="assembly\_fragment:01051"  
fragment\_chain:2  
97875..103188  
/note="assembly\_fragment:00803"  
fragment\_chain:2  
103289..117057

/note="assembly\_fragment:00957"  
fragment\_chain:2  
117158..121704  
/note="assembly\_fragment:01825"  
fragment\_chain:2  
121805..123958  
/note="assembly\_fragment:00145"  
fragment\_chain:2  
124059..129614  
/note="assembly\_fragment:01306"  
fragment\_chain:3  
129715..132913  
/note="assembly\_fragment:01042"  
fragment\_chain:3  
133014..141267  
/note="assembly\_fragment:00839"  
fragment\_chain:3  
141368..147593  
/note="assembly\_fragment:00193"  
fragment\_chain:3  
147694..159744  
/note="assembly\_fragment:00848"  
fragment\_chain:3  
159845..170736  
/note="assembly\_fragment:01745"  
fragment\_chain:3  
170837..173576  
/note="assembly\_fragment:00061"  
173677..185245  
/note="assembly\_fragment:00941"  
185346..188201  
/note="assembly\_fragment:01166"  
clone\_end:17  
vector\_side:right  
BASE COUNT 54379 a 40690 c 39102 g 51726 t 2304 others  
ORIGIN  
Query Match 49.2%; Score 25.6; DB 2; Length 188201;  
Best Local Similarity 77.5%; Pred. No. 19;  
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 13 AAAGTTCATGCTTTAGGCCCAAACTAGGTCAAAGGTCA 52  
|||||  
DB 69280 AGAGACATGTCATTACACCAAACTAGGCAAGTTCA 69319  
AC128896 172453 bp DNA linear HTG 24-JUL-2002  
Rattus norvegicus clone CH230-192P7, \*\*\* SEQUENCING IN PROGRESS  
\*\*\* 56 unordered pieces.  
AC128896 AC128896.1 GI:21953296  
HTG; HTGS\_PHASE1.  
Rattus norvegicus.  
Rattus norvegicus.  
Rattus norvegicus.  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 172453)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T.,  
Barbieri, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, M., Brown, N.P.,  
Buhay, C., Burch, P., Burrell, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.B., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,



Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Hawley, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
 Homs, F., Howard, S., Huber, J., Hui, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, J., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhinney, B., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogum, M., Okwuonu, G.,  
 Oragunye, N., Oviedo, R., Primus, E., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Pimentel, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Sparks, A., Stanley, H., Stone, H.,  
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telifrod, B., Thomas, N., Thomas, S.,  
 Umari, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wlarczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Center: Baylor College of Medicine  
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: KBGD  
 Center clone name: CH230-192P7  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap; version 0.990323  
 Consensus quality: 114919 bases at least Q40  
 Consensus quality: 121874 bases at least Q30  
 Consensus quality: 127181 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence; it currently  
 \* consists of 56 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1105: contig of 1105 bp in length  
 \* 1106: gap of unknown length  
 \* 1206: contig of 1669 bp in length  
 \* 2874: gap of unknown length  
 \* 2975: contig of 1348 bp in length  
 \* 4322: gap of unknown length  
 \* 4323: contig of 1428 bp in length  
 \* 4423: gap of unknown length  
 \* 5850: contig of 1130 bp in length  
 \* 5951: gap of unknown length  
 \* 7080: contig of 1130 bp in length  
 \* 7180: gap of unknown length  
 \* 7181: contig of 1264 bp in length  
 \* 8445: gap of unknown length  
 \*  
 \* 8545: contig of 1144 bp in length  
 \* 9689: gap of unknown length  
 \* 9789: contig of 1028 bp in length  
 \* 10817: gap of unknown length  
 \* 12361: contig of 1445 bp in length  
 \* 12461: gap of unknown length  
 \* 13612: contig of 1151 bp in length  
 \* 13712: gap of unknown length  
 \* 14771: contig of 1059 bp in length  
 \* 14871: gap of unknown length  
 \* 16387: contig of 1516 bp in length  
 \* 16388: gap of unknown length  
 \* 17908: contig of 1421 bp in length  
 \* 18008: gap of unknown length  
 \* 19035: contig of 1027 bp in length  
 \* 19135: gap of unknown length  
 \* 20283: contig of 1148 bp in length  
 \* 20383: gap of unknown length  
 \* 21766: contig of 1793 bp in length  
 \* 22776: gap of unknown length  
 \* 24132: contig of 1856 bp in length  
 \* 24232: gap of unknown length  
 \* 26007: contig of 1775 bp in length  
 \* 26107: gap of unknown length  
 \* 28158: contig of 2051 bp in length  
 \* 28258: gap of unknown length  
 \* 31278: contig of 3020 bp in length  
 \* 31378: gap of unknown length  
 \* 33780: contig of 2402 bp in length  
 \* 33800: gap of unknown length  
 \* 35306: contig of 1426 bp in length  
 \* 35406: gap of unknown length  
 \* 37823: contig of 2417 bp in length  
 \* 37924: gap of unknown length  
 \* 39812: contig of 1889 bp in length  
 \* 39813: gap of unknown length  
 \* 39812: gap of unknown length  
 \* 41980: contig of 2068 bp in length  
 \* 42080: gap of unknown length  
 \* 43623: contig of 1543 bp in length  
 \* 43723: gap of unknown length  
 \* 45037: contig of 1314 bp in length  
 \* 45137: gap of unknown length  
 \* 48313: contig of 3176 bp in length  
 \* 48413: gap of unknown length  
 \* 50357: contig of 1944 bp in length  
 \* 50457: gap of unknown length  
 \* 53773: contig of 3316 bp in length  
 \* 53873: gap of unknown length  
 \* 56643: contig of 2770 bp in length  
 \* 56743: gap of unknown length  
 \* 58788: contig of 2045 bp in length  
 \* 58888: gap of unknown length  
 \* 61905: contig of 3017 bp in length  
 \* 62005: gap of unknown length  
 \* 64472: contig of 2467 bp in length  
 \* 64572: gap of unknown length  
 \* 67206: contig of 2634 bp in length  
 \* 67306: gap of unknown length  
 \* 69999: contig of 2693 bp in length  
 \* 70099: gap of unknown length  
 \* 71592: contig of 1493 bp in length  
 \* 71593: gap of unknown length  
 \* 71692: gap of unknown length  
 \* 75202: contig of 3510 bp in length  
 \* 75203: gap of unknown length  
 \* 75302: gap of unknown length  
 \* 77610: gap of unknown length  
 \* 77710: gap of unknown length  
 \* 80648: contig of 2938 bp in length  
 \* 80748: gap of unknown length  
 \* 84518: contig of 3770 bp in length  
 \* 84618: gap of unknown length  
 \* 84519: gap of unknown length  
 \* 84619: contig of 2457 bp in length  
 \* 87175: gap of unknown length  
 \* 91567: contig of 4392 bp in length

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

```

* 91568 91667: gap of unknown length
* 91668 96457: contig of 4790 bp in length
* 96458 96557: gap of unknown length
* 96557 99936: contig of 3379 bp in length
* 99937 100036: gap of unknown length
* 100037 103459: contig of 3423 bp in length
* 103460 103559: gap of unknown length
* 103560 107728: contig of 4169 bp in length
* 107729 107828: gap of unknown length
* 107829 113390: contig of 5562 bp in length
* 113391 113490: gap of unknown length
* 113491 116984: contig of 5494 bp in length
* 116985 119084: gap of unknown length
* 119085 124631: contig of 5547 bp in length
* 124632 124731: gap of unknown length
* 124732 130602: contig of 5871 bp in length
* 130603 130702: gap of unknown length
* 130703 136270: contig of 5568 bp in length
* 136271 136370: gap of unknown length
* 136371 145489: contig of 9119 bp in length
* 145490 145589: gap of unknown length
* 145590 152194: contig of 6605 bp in length
* 152195 152294: gap of unknown length
* 152295 160898: contig of 8604 bp in length
* 160899 160998: gap of unknown length
* 160999 172453: contig of 11455 bp in length.

FEATURES             Location/Qualifiers
     source            1..172453
                        /organism="Rattus norvegicus"

Query Match          48.8%; Score 25.4; DB 2; Length 172453;
Best Local Similarity 68.6%; Pred. No. 23;
Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

```

2y 2 AAAAAGGTCAGGTCATGCTTTAGGCCCCAAAAGTCAAGTCA 52
|||||
Db 156579 AAAAAGGTCAGGTCATGCTTTAGGCCCCAAAAGTCAAGTCA 156529
|||||

```

```

RESULT 13
AC025286
LOCUS Homo sapiens chromosome 16 clone RP11-478M13, WORKING DRAFT
DEFINITION AC025286 127286 bp DNA linear HTG 18-JUL-2000
SEQUENCE AC025286.3 GI:3256469
ACCESSION AC025286
VERSION HTG; HTGS PHASE2; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 127286)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
2 (bases 1 to 127286)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (08-MAR-2000) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7705027.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: RP11-478M13
Center clone name: RP11-478M13
-----
Summary Statistics
Consensus quality: 124330 bases at least Q40
Consensus quality: 126704 bases at least Q30

```

```

Consensus quality: 127028 bases at least Q20
Estimated insert size: 147150; agarose-fp estimation
Estimated insert size: 127286; sum-of-ctngs estimation
Quality coverage: 7.14 in Q20 bases; agarose-fp estimation
Quality coverage: 8.25 in Q20 bases; sum-of-ctngs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 50500: contig of 50500 bp in length
* 50501 127286: contig of 76686 bp in length.

FEATURES             Location/Qualifiers
     source            1..127286
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosomes="16"
                        /clone="RP11-478M13"
                        /clone_lib="RP11-478M13"
                        /clone_lib="RP11-478M13"
BASE COUNT 30667 a 30615 c 31390 g 34514 t 100 others
ORIGIN
Query Match          48.5%; Score 25.2; DB 2; Length 127286;
Best Local Similarity 78.9%; Pred. No. 28;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 CAAAACCTAGGTCAAGGTCATGCTTTAGGCCCCAAAAC 38
Db 69472 CAAAACCTAGGTCAAGGTCATGCTTTAGGCCCCAAAAC 69509

RESULT 14
AC117282
LOCUS Rattus norvegicus clone CH230-358117, *** SEQUENCING IN PROGRESS
DEFINITION AC117282 155491 bp DNA linear HTG 17-JUL-2002
SEQUENCE AC117282.2 GI:21746067
KEYWORDS HTG; HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 155491)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshan, F.R., Allen, C.,
Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbosa, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsberg, E., Kelly, S., Khan, J., King, J., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisedge, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,

```

Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, G., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwona, G.,  
 Oragunye, N., Oviator, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, J., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, G., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sison, I.,  
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vaquez, B., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 155491)  
 Worley, K.C.  
 Direct Submission  
 Submitted (10-APR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 155491)  
 Worley, K.C.  
 Direct Submission  
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 14, 2002 this sequence version replaced gi:20127206.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GTEJ  
 Center clone name: CH230-358117  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 115768 bases at least Q40  
 Consensus quality: 121766 bases at least Q30  
 Consensus quality: 125805 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 40 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1477: contig of 1477 bp in length  
 \* 1478 1577: gap of unknown length  
 \* 1578 2222: contig of 1345 bp in length  
 \* 2923 3022: gap of unknown length  
 \* 3023 4036: contig of 1014 bp in length  
 \* 4037 4136: gap of unknown length  
 \* 4137 5570: contig of 1434 bp in length  
 \* 5571 5671: gap of unknown length  
 \* 5671 7006: contig of 1336 bp in length  
 \* 7007 7106: gap of unknown length  
 \* 7107 8615: contig of 1509 bp in length  
 \* 8616 10126: contig of 1411 bp in length  
 \* 8716 10127: gap of unknown length  
 \* 10127 10226: gap of unknown length  
 \* 10227 11498: contig of 1272 bp in length

# JOURNAL REFERENCE AUTHORS JOURNAL JOURNAL REFERENCE AUTHORS JOURNAL JOURNAL COMMENT

11499 11598: gap of unknown length  
 11599 12923: contig of 1325 bp in length  
 12923 13023: gap of unknown length  
 13024 14740: contig of 1717 bp in length  
 14741 14840: gap of unknown length  
 14841 16538: contig of 1698 bp in length  
 16539 16638: gap of unknown length  
 16639 19845: contig of 3207 bp in length  
 19846 19945: gap of unknown length  
 19946 21394: contig of 1449 bp in length  
 21395 21494: gap of unknown length  
 21495 23018: contig of 1524 bp in length  
 23019 23118: gap of unknown length  
 23119 24864: contig of 1746 bp in length  
 24865 24964: gap of unknown length  
 24965 28816: contig of 3852 bp in length  
 28817 31066: contig of 2150 bp in length  
 31067 31166: gap of unknown length  
 31167 34835: contig of 3669 bp in length  
 34836 34935: gap of unknown length  
 34936 37513: contig of 2577 bp in length  
 37513 37613: gap of unknown length  
 37613 40018: contig of 2406 bp in length  
 40019 40118: gap of unknown length  
 40119 41661: contig of 1543 bp in length  
 41662 41761: gap of unknown length  
 41762 44947: contig of 3186 bp in length  
 44948 45047: gap of unknown length  
 45048 47637: contig of 2590 bp in length  
 47638 47737: gap of unknown length  
 47738 52326: contig of 4589 bp in length  
 52327 52426: gap of unknown length  
 52427 54974: contig of 2548 bp in length  
 54975 55074: gap of unknown length  
 55075 58895: contig of 3821 bp in length  
 58896 58995: gap of unknown length  
 58996 62384: contig of 3389 bp in length  
 62385 62484: gap of unknown length  
 62485 65811: contig of 3326 bp in length  
 65812 70912: gap of unknown length  
 70913 71012: contig of 5002 bp in length  
 71013 74308: contig of 3296 bp in length  
 74309 74408: gap of unknown length  
 74409 79274: contig of 4866 bp in length  
 79275 79374: gap of unknown length  
 79375 84900: contig of 5526 bp in length  
 84901 85000: gap of unknown length  
 85001 89777: contig of 4777 bp in length  
 89778 96801: contig of 6924 bp in length  
 96802 96901: gap of unknown length  
 96902 103026: contig of 6125 bp in length  
 103027 103126: gap of unknown length  
 103127 109834: contig of 6708 bp in length  
 109835 121230: gap of unknown length  
 121231 121330: contig of 11296 bp in length  
 121331 131715: contig of 10385 bp in length  
 131716 131815: gap of unknown length  
 131816 142666: contig of 10851 bp in length  
 142667 142766: gap of unknown length  
 142767 155491: contig of 12725 bp in length.  
 Location/Qualifiers  
 1. 155491  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-358117"  
 BASE COUNT 41193 a 33783 c 33209 g 41856 t 5450 others  
 ORIGIN  
 Query Match 48.5%; Score 25.2; DB 2; Length 155491;

Job time : 518.727 secs

Best Local Similarity 71.7%; Pred. No. 27; Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

2 AAAACTAGGTCAAAGGTCATGCTTTTAGGCCCAAACTAGGTCAA 47  
db 32757 AAGTCAAACTCAAGGTCATGCTTTTAGGCCCTTGCTAGTTCAAA 32802

RESULT 15  
AC092723/c  
.LOCUS AC092723 189458 bp DNA linear PRI 06-FEB-2002  
DEFINITION Homo sapiens chromosome 16 clone RP11-542M13, complete sequence.  
ACCESSION AC092723 AC068615  
VERSION AC092723.3 GI:18542980  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 189458)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
UNPUBLISHED  
2 (bases 1 to 189458)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (21-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 189458)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 6, 2002 this sequence version replaced gi:15055296.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

FEATURES  
source Location/Qualifiers  
1..189458  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-542M13"  
BASE COUNT 49597 a 48044 c 45725 g 45092 t  
ORIGIN

Query Match 48.5%; Score 25.2; DB 9; Length 189458;  
Best Local Similarity 78.9%; Pred. No. 27;  
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAAGGTCATGCTTTTAGGCCCAAAAC 38  
Db 180206 CAAATGGGTCAAAGTTCATGCTTTTAGGGCAGATAC 180169

Search completed: February 19, 2003, 23:06:36

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

run on: February 19, 2003, 19:26:38 ; Search time 31.6042 Seconds  
(without alignments)  
3705.323 Million cell updates/sec

Title: US-09-808-388-4  
Perfect score: 52  
Sequence: 1 caaaactaggtcaaggtca.....caaaactaggtcaaaaggtca 52

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	52	22	AAI64306
2	24.8	47.7	408	21	AAA50275
3	23.6	45.4	3674	13	AAQ29114
C 4	23.2	44.6	1681	8	AAV70720
C 5	23	44.2	648	23	ABV05301
C 6	23	44.2	867	24	ABN98803
C 7	22.8	43.8	400	22	AAH05944
C 8	22.8	43.8	2537	22	AAH17687
9	22.4	43.1	851	22	AAI94417

C 10	22.2	42.7	51	22	AAI27493
C 11	22.2	42.7	2019	20	AAI61766
C 12	22.2	42.7	2154	20	AAI61765
C 13	22.2	42.7	2716	11	AAQ02831
C 14	22.2	42.7	111309	20	AAI20250
15	22	42.3	1168	21	AAI39133
16	21.8	41.9	352	24	ABQ85811
17	21.8	41.9	440	21	AAI05939
18	21.8	41.9	441	23	ABV53997
19	21.8	41.9	1314	21	AAI37749
20	21.8	41.9	1506	24	AAI37066
C 21	21.8	41.9	2086	22	AAI37066
C 22	21.8	41.9	2214	22	AAI65995
C 23	21.8	41.9	6027	21	AAI13061
C 24	21.8	41.9	12810	21	AAI95272
C 25	21.8	41.9	72928	20	AAI18355
C 26	21.8	41.9	72928	21	AAI80253
C 27	21.4	41.2	379	22	AAI90897
C 28	21.4	41.2	1206	22	AAI94515
C 29	21.4	41.2	349980	22	AAI41226
C 30	21.2	40.8	428	22	AAI33634
C 31	21.2	40.8	653	24	ABQ57345
C 32	21.2	40.8	787	20	AAI15615
C 33	21.2	40.8	873	23	AAI65160
C 34	21.2	40.8	924	23	AAI51621
C 35	21.2	40.8	960	23	AAI54377
C 36	21.2	40.8	1008	24	AAI30356
C 37	21.2	40.8	1482	22	AAI33428
C 38	21.2	40.8	1943	23	AAI93825
C 39	21.2	40.8	3135	18	AAI74772
C 40	21.2	40.8	10864	22	AAI46262
C 41	21.2	40.8	73308	24	ABL65966
C 42	21	40.4	38	24	ABK19742
C 43	21	40.4	41	22	AAI64305
C 44	21	40.4	311	23	ABV35630
C 45	21	40.4	332	22	AAI64308

ALIGNMENTS

RESULT 1  
AAI64306  
ID AAI64306 standard; DNA; 52 BP.  
XX AAI64306;  
AC AAI64306;  
DT 15-NOV-2001 (first entry)  
XX PPAR response element (DR1) 2 31.  
DE PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2s11A;  
KW peroxisome proliferator activated receptor;  
KW secreted non-pancreatic phospholipase A2; ss.  
XX Synthetic.  
XX WO200168845-A2.  
XX 20-SEP-2001.  
XX 14-MAR-2001; 2001WO-FR00759.  
XX 14-MAR-2000; 2000FR-0003262.  
XX 13-APR-2000; 2000US-0196959.  
XX (AVET) AVENTIS PHARMA SA.  
XX Massaad C, Berenbaum F, Olivier J, Salvat C, Berezat G;  
XX WPI; 2001-582451/65.

PT New hybrid promoter induced by inflammation, useful in gene therapy of  
 PT arthritis, comprises peroxisome proliferator activated receptor  
 PT response element and promoter of secreted phospholipase A2

PS Claim 4; Page 29; 52pp; French.

XX  
 CC The present invention relates to a hybrid promoter comprising (i) a PPAR  
 CC (peroxisome proliferator activated receptor) response element (PPRE); and  
 CC (ii) at least part of the promoter of the PLA2SIIA (secreted  
 CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to  
 CC regulate expression of therapeutic transgenes, for experimental,  
 CC clinical, therapeutic or diagnostic use, especially in chondrocytes for  
 CC treatment of arthritis, but also in bone, muscle, liver, heart, the  
 CC nervous system and tumours. The present sequence is a PPAR response  
 CC element, which was used to generate the hybrid promoter of the present  
 CC invention.

SQ Sequence 52 BP; 19 A; 11 C; 11 G; 11 T; 0 other;

Query Match 100.0%; Score 52; DB 22; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-11;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAGAGGTCAGGTCATGCTTTAGGCCCAAACTAGGTCAGGTCACA 52

Db 1 CAAGAGGTCAGGTCATGCTTTAGGCCCAAACTAGGTCAGGTCACA 52

RESULT 2

AAA50275

ID AAA50275 standard; DNA; 408 BP.

XX AC AAA50275;

XX 07-NOV-2000 (first entry)

DT Human LSR gene 5' regulatory region fragment.

DE Lipolysis stimulated receptor; LSR; chromosome 19q13.1; human;  
 KW single nucleotide polymorphism; biallelic marker; obesity;  
 KW atherosclerosis; insulin resistance; hypertension; hyperlipidemia;  
 KW hypertriglyceridemia; cardiovascular disease; microangiopathy;  
 KW syndrome X; diagnosis; therapy; genotyping; ds.

XX Homo sapiens.

XX Key Location/Qualifiers  
 EH replace(353,A)

FT /\*tag= a

FT /standard name= "single nucleotide polymorphism"

FT /note= "marker 99-14424-353 (A2)"

FT complement (1..20)

FT /\*tag= b

FT /note= "upstream amplification primer"

FT 388..408

FT /\*tag= c

FT /note= "downstream amplification primer"

XX WO200047772-A2.

XX 17-AUG-2000.

XX 08-FEB-2000; 2000WO-1800194.

XX 10-FEB-1999; 99US-0119592.

XX 20-JUL-1999; 99US-0144784.

XX (GEST ) GENSET.

XX Blumenfeld M, Bougueleret L, Bihain B;

XX WPI; 2000-506098/45.

PT New isolated or recombinant LSR polynucleotide useful for genotyping  
 PT and detecting human risk of developing detectable trait e.g. obesity,  
 PT comprises LSR-related biallelic marker

PS Claim 1; Page 165; 191pp; English.

XX  
 CC The present sequence represents a portion of the 5' regulatory  
 CC regions of the human lipolysis stimulated receptor (LSR) gene. The  
 CC sequence was obtained by PCR amplification of genomic DNA. It  
 CC includes biallelic marker (single nucleotide polymorphism) A2, which  
 CC can be detected by methods of the invention. LSR is involved in  
 CC the partitioning of dietary lipids between the liver and peripheral  
 CC tissues, including adipose tissue. The invention is directed to  
 CC biallelic markers located within the LSR genomic sequence (see also  
 CC AAA50273-84). The markers provide useful tools for identifying  
 CC associations between specific alleles of the LSR gene and obesity  
 CC or a related disorder, such as atherosclerosis, insulin resistance,  
 CC hypertension, hyperlipidemia, hypertriglyceridemia, cardiovascular  
 CC disease, microangiopathy in obese individuals with Type II diabetes,  
 CC ocular and renal lesions associated with microangiopathy in such  
 CC individuals, and Syndrome X. The biallelic markers of the LSR  
 CC gene can lead to the identification of new targets acting against  
 CC obesity or obesity-related disorders. They can also be used to  
 CC diagnose a susceptibility to obesity or to identify the cause of  
 CC obesity for an individual. Preferred biallelic markers are  
 CC selected from A1-A32 and A1-A20, especially A2, A15, A16, A17,  
 CC A21, A23, A24, A26 and A31, particularly A15, A17 and A21. Primers  
 CC hybridizing to regions flanking the biallelic markers are provided,  
 CC as well as methods for genotyping a nucleic acid for 1 or more  
 CC biallelic markers, and for detecting a statistical correlation  
 CC between a biallelic marker allele and a phenotype and/or between a  
 CC biallelic marker haplotype and a phenotype.

XX SQ Sequence 408 BP; 119 A; 107 C; 90 G; 89 T; 3 other;

Query Match 47.7%; Score 24.8; DB 21; Length 408;

Best Local Similarity 80.6%; Pred. No. 2.5;

Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 12 CAAGGTCATGCTTTAGGCCCAAACTAGGTCACAA 47

Db 348 CACACGACATGGCTTTAGGCCCAAACTAGGTCACAA 383

RESULT 3

AAQ29114

ID AAQ29114 standard; DNA; 3674 BP.

XX AC AAQ29114;

XX 24-FEB-1993 (first entry)

XX FeLV-A gag/pol fragment.

XX AIVAC; feline leukemia virus; FeLV-A; gag; pol; pFGA-2 gag; PCR;  
 XX polymerase chain reaction; vaccinia virus; H6 promoter; PC3FGAG;  
 XX PC3FGAGVQ; expression cassette; PC3DOFGAGVQ; vCP83; vCP87; ATI;  
 XX rescue virus; vCP97; immunosuppressive region; vCP93; polylinker;  
 XX A type inclusion body region; PSD541; ss.

XX Synthetic.

XX Key Location/Qualifiers

FT misc\_feature 652..654

FT /\*tag= a

FT /label= Initiation\_codon

XX WO9215672-A.

XX 17-SEP-1992.

XX 09-MAR-1992; 92WO-US01906.

PR 07-MAR-1991; 91US-0666056.  
 PR 11-JUN-1991; 91US-0713967.  
 PR 06-MAR-1992; 92US-0847951.  
 XX  
 PA (VIRO-) VIROGENETICS CORP.  
 XX  
 PI Cox WI, De Taisne C, Francis J, Gettig RR, Johnson GP;  
 PI Limbach KJ, Norton EK, Paoletti E, Perkus ME, Pincus SE;  
 PI Riviere M, Tartaglia J, Taylor J;  
 XX  
 DR WPI, 1992-331718/40.  
 XX  
 PT Vaccine comprises recombinant, attenuated pox-virus - use for  
 PT vaccinating against viral infections such as rabies, hepatitis B,  
 PT HIV, HSV, EBV, CMV, mumps etc.  
 XX  
 PS Disclosure; Fig 27; 456pp; English.  
 XX  
 CC The sequence given encodes the feline leukemia virus (FeLV-A) gag  
 CC gene. This sequence was used in the construction of an ALVAC-based  
 CC recombinant containing the FeLV-A gag gene. The gag/pol sequence was  
 CC derived from plasmid pFGA-2 gag. The gag gene was liberated on a 2.5  
 CC kb fragment and was amplified by polymerase chain reaction (PCR).  
 CC This sequence was aligned with the vaccinia virus H6 promoter. The  
 CC FeLV-A pol gene was also included in this construction to yield a  
 CC plasmid designated pC3PGAG. Termination codons were included to form  
 CC the plasmid pC3PGAGVQ. The FeLV gag/pol expression cassette was  
 CC excised from this plasmid and designated pC3DORGAGVQ. pC3DORGAGVQ  
 CC was used in standard in vitro recombination assays with vCP83 and  
 CC vCP87 as rescue viruses. Recombinants containing the entire FeLV-A  
 CC gag/pol sequences and the entire FeLV-A env gene were designated  
 CC vCP97, while recombinants containing the same gag/pol sequences and  
 CC the entire FeLV-A env lacking the immuno-suppressive region were  
 CC designated vCP93. The FeLV-A gag can be inserted into a vaccinia  
 CC virus background by including a polylinker sequence flanking the  
 CC coding region and using these sequences to insert the gag region into  
 CC the A type inclusion body region (ARI) of the plasmid pSD541. See  
 CC also AAQ35501-864.  
 XX  
 SQ Sequence 3674 BP; 970 A; 1018 C; 883 G; 803 T; 0 other;  
 Query Match 45.4%; Score 23.6; DB 13; Length 3674;  
 Best Local Similarity 69.6%; Pred. No. 13;  
 Matches 32; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 XX  
 QY 4 AACTAGTCAAGGTCATCTTTAGGCCCAAACTAGTCAAGG 49  
 Db 1661 AGAAAGGCAAGGGTCCTTCTGAGGCCCAAACTAGTCCAGG 1706  
 XX  
 RESULT 4  
 AAN70720/c  
 ID AAN70720 standard; cDNA; 1681 BP.  
 XX  
 AC AAN70720;  
 XX  
 DT 08-MAR-1992 (first entry)  
 XX  
 DE Sequence of mcf3 cDNA which encodes the ros 1 oncogene.  
 XX  
 KW Activated oncogene; diagnosis; tumour; therapy; predisposition;  
 KW astrocytoma; glioblastoma; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer  
 XX

PF 21-MAY-1987; 87WO-US01215.  
 XX  
 PR 19-MAY-1987; 87US-0048915.  
 PR 21-MAY-1986; 86US-0865569.  
 XX  
 PA (COLD-) COLD SPRING HARBOR.  
 XX  
 PI Wigler MH, Birchmeier C, Fasano O, Birnbaum D;  
 XX  
 DR WPI; 1987-348652/49.  
 DR P-PSDE; AAP70443.  
 XX  
 PT DNA sequence encoding human ros oncogene - used in detection of  
 PT tumour cells and pre-disposition to disease and in tumour  
 PT treatment  
 XX  
 PS Claim 1; Fig 1; 43pp; English.  
 XX  
 CC The claimed human ros 1 oncogene (AAN70720) encodes a polypeptide  
 CC capable of transforming NIH3T3 cells and of inducing a tumour when  
 CC injected into nude mice. The polypeptide, which is also claimed  
 CC (AAP70443) has tyrosine-specific protein kinase activity. AAN70719 was  
 CC used as a probe to localise exon 5's of mcf3 cDNA.  
 XX  
 SQ Sequence 1681 BP; 504 A; 314 C; 407 G; 456 T; 0 other;  
 Query Match 44.6%; Score 23.2; DB 8; Length 1681;  
 Best Local Similarity 70.5%; Pred. No. 16;  
 Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 XX  
 QY 2 AAACTAGTCAAGGTCATCTTTAGGCCCAAACTAGTCA 45  
 Db 1067 AAAAGTAGTCTTTGGTGGGCTTCTGAGGCCCACTGGGTCA 1024  
 XX  
 RESULT 5  
 ABV05301/c  
 ID ABV05301 standard; cDNA; 648 BP.  
 XX  
 AC ABV05301;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 5292.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US05171.  
 XX  
 PR 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer  
 XX

Claim 1; Page 897; 11750pp; English.

PS The invention relates to an isolated nucleic acid molecule (I) comprising  
 XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 648 BP; 154 A; 153 C; 147 G; 138 T; 56 other;

Query Match 44.2%; Score 23; DB 23; Length 648;  
 Best Local Similarity 67.4%; Pred. No. 15;  
 Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 5 ACTAGGTCARAGGTCAGTCTTTAGGCCCAAACTAGGTCAAA 47  
 Db 361 AATTGGAAAAAGGTTTGGCINTANCCNCAANTTGGTCAGA 319

# RESULT 6

ABN98803

ID ABN98803 standard; DNA; 867 BP.

XX AC ABN98803;

XX 01-AUG-2002 (first entry)

DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 571.

XX Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;  
 KW disease; crop; thale cress; tolerance factor; insect; pathogen;  
 KW nutrition; ds.

XX Arabidopsis thaliana.

XX US2002023281-A1.

XX 21-FEB-2002.

XX 26-JAN-2001; 2001US-0770445.

XX 27-JAN-2000; 2000US-178472P.

XX (GORLACH J.

PA (ANYI/) AN Y.

PA (HAMI/) HAMILTON C M.

PA (PRIC/) PRICE J L.

PA (RAIN/) RAINES T M.

PA (YUYI/) YU Y.

PA (RAME/) RAMEAKA J G.

PA (PAGE/) PAGE A.

PA (MATH/) MATHEN A V.

PA (LEDF/) LEDFORD B L.

PA (WOES/) WOESSNER J P.

PA (HAAS/) HAAS W D.

PA (GARC/) GARCIA C A.

PA (KRICK/) KRICKER M.

PA (SLAT/) SLATER T.

PA (DAVI/) DAVIS K R.

PA (ALLE/) ALLEN K.

PA (HOFF/) HOFFMAN N.

PA (HURE/) HURBAN P.

XX

PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,  
 PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;  
 PI Garcia CA, Kricker M, Slater T, Davis K, Allen K, Hoffman N;  
 PI Hurban P;

XX WPI; 2002-400781/43.

XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,  
 PT producing compositions that modulate the expression or function of its  
 PT encoded protein, and mapping functional regions of protein -

XX Claim 1; SEQ ID NO 571; 49pp + Sequence Listing; English.

XX The invention relates to an Arabidopsis thaliana nucleic acid (I)  
 CC comprising a sequence capable of hybridising under stringent conditions  
 CC to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),  
 CC given in the specification or its fragment. A polypeptide (II) encoded by  
 CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a  
 CC genetically modified cell (IV) comprising an exogenous nucleic acid, is  
 CC useful for screening a candidate agent for its biological effect. (I) is  
 CC useful in identifying homologous or related genes, in producing  
 CC compositions that modulate the expression or function of its encoded  
 CC protein, mapping functional regions of the protein and in studying  
 CC associated physiological pathways. (I) is also useful for the genetic  
 CC manipulation of cells, particularly plant cells. (I) is also useful in  
 CC screening assays of various plant strains to determine the strains that  
 CC are best capable of withstanding a particular disease or environmental  
 CC stress. (II) and (III) are useful for screening of biologically active  
 CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical  
 CC pathways. The screened agents are useful in improved methods of treating  
 CC crops to prevent or treat disease. (II) are also useful in screening  
 CC programs to identify agents that mimic or enhance the action of tolerance  
 CC factors. Such agents are useful in improved methods of treating crops to  
 CC enhance their tolerance to environmental stress. (I) is also useful  
 CC for enhancing or inhibiting production of a biosynthetic product in a  
 CC plant. (III) is useful for identifying other mediators that may induce  
 CC expression of proteins of interest, for establishing the extent to which  
 CC any specific insect and/or pathogen is responsible for damage to a  
 CC particular plant, for identifying other mediators that enhance or induce  
 CC tolerance to environmental stress, for identifying factors involved in  
 CC biosynthetic pathways of nutritional, commercial, or medicinal value and  
 CC for identifying productions of nutritional, commercial or medicinal  
 CC value. (IV) is useful in the study of genetic function and regulation,  
 CC for alteration of the cellular metabolism and for screening compounds  
 CC that may affect the biological function of the gene or gene products.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?DocID=99909770445.

XX Sequence 867 BP; 247 A; 205 C; 168 G; 244 T; 3 other;

Query Match 44.2%; Score 23; DB 24; Length 867;

Best Local Similarity 70.7%; Pred. No. 16;

Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 7 TAGGTCAAAGGTCATGCTTTAGGCCCAAACTAGGTCAAA 47

Db 606 TAGGTCAACATCAGTCTTTGGCGCACACTAGATNNAA 646

# RESULT 7

AAH05944/c

ID AAH05944 standard; cDNA; 400 BP.

XX AC AAH05944;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:2779.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

OS





```

XX OS Homo sapiens.
XX PI WO200166719-A1.
XX FN
XX XX
XX PD 13-SEP-2001.
XX PF 02-MAR-2001; 2001WO-JF01629.
XX XX
XX PD 07-MAR-2000; 2000JP-0159195.
XX PF
XX PA (CHIB-) CHIBA PREFECTURE.
XX PA (HISM) HISAMITSU PHARM CO LTD.
XX PI Nakagawara A;
XX DR WPI; 2001-565584/63.
XX XX
XX XX Nucleic acids originating in gene expressed in human neuroblastoma,
XX XX useful as probe or primer in diagnosing prognosis of human
XX XX neuroblastoma, malignancy and susceptibility indicator or tumour marker
XX XX for anti-cancer agents.
XX XX
XX XX Claim 1; Page 406-407; 2979pp; Japanese.
XX XX
XX XX The invention relates to novel genes (AA193926-AA197963) expressed in
XX XX human neuroblastoma. The nucleic acids are applicable as a probe or
XX XX primer in diagnosing the prognosis of human neuroblastoma, malignancy and
XX XX susceptibility indicators or tumour markers for anti-cancer agents. The
XX XX gene information for diagnosing prognosis is related to factors similar
XX XX to that for N-myc and TrkA genes.
XX XX
XX XX Sequence 851 BP; 268 A; 160 C; 119 G; 262 T; 42 other;
XX XX
XX XX Query Match 43.1%; Score 22.4; DB 22; Length 851;
XX XX Best Local Similarity 72.5%; Pred. No. 28;
XX XX Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX XX
XX QY 11 TCAAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGGT 50
XX DB 192 TGAGAGGTCAAGTCTTTAGGCAGGAAACAGGCAGAAAGGT 231
XX
XX RESULT 10
XX AAL27493/c
XX ID AAL27493 standard; DNA; 51 BP.
XX AC
XX AC AAL27493;
XX DT
XX DT 24-JAN-2002 (first entry)
XX XX
XX DE Human SNP oligonucleotide #701.
XX XX
XX XX Immunosuppressive; immunostimulatory; antinflammatory; cytostatic;
XX KW neuroprotective; antimicrobial; gene therapy; vaccine; amylose; cancer;
XX KW amyloid protein; angiotensin; apoptosis related protein; cadherin;
XX KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
XX KW complement related protein; cytochrome; kinesis; cytokine; interferon;
XX KW interleukin; G-protein coupled receptor; thioesterase; inflammation;
XX KW multifactorial disease; autoimmune disease; infection;
XX KW nervous system disease; ss.
XX OS
XX OS Homo sapiens.
XX XX
XX XX WO200147944-A2.
XX XX
XX XX 05-JUL-2001.
XX XX
XX XX 28-DEC-2000; 2000WO-US35498.
XX PF
XX PF 28-DEC-1999; 99US-0173419.
XX PR
XX PR 27-DEC-2000; 2000US-0173419.
XX XX

```

```

PA (CURA-) CURAGEN CORP.
XX XX
XX PI Shimkets RA, Leach M;
XX XX
XX DR WPI; 2001-465210/50.
XX XX
XX XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
XX XX oncogenes and histones, useful for diagnosing and treating, e.g.
XX XX cancer, autoimmune diseases and infections.
XX XX
XX XX Claim 1; Page 1586; 4143pp; English.
XX XX
XX XX The present invention relates to oligonucleotides encoding polymorphic
XX XX variants of proteins related to amylases, amyloid proteins, angiotensin,
XX XX apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,
XX XX histones, kinases, colony stimulating factors, complement related
XX XX proteins, cytochromes, kinesis, cytokines, interferons, interleukins,
XX XX G-protein coupled receptors and thioesterases. The present sequence is
XX XX one such oligonucleotide. The oligonucleotides and the peptides encoded
XX XX by them may be used in the prevention, diagnosis and treatment of
XX XX diseases associated with inappropriate expression of the proteins listed
XX XX above. Disorders that may be prevented, diagnosed and/or treated include
XX XX multifactorial diseases with a genetic component, such as autoimmune
XX XX diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,
XX XX systemic lupus erythematosus and Grave's disease), inflammation, cancer
XX XX (e.g. cancers of the bladder, brain, breast, colon and kidney,
XX XX CC leukaemia), diseases of the nervous system and an infection of pathogenic
XX XX organisms.
XX XX
XX XX Sequence 51 BP; 10 A; 17 C; 10 G; 14 T; 0 other;
XX XX
XX XX Query Match 42.7%; Score 22.2; DB 22; Length 51;
XX XX Best Local Similarity 69.8%; Pred. No. 17;
XX XX Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
XX XX
XX QY 3 AAATAGGTCAAGGTCATGCTTTAGGCCCAAACTAGGTCA 45
XX DB 51 AAATAGGACGAGTCCAGTGTGTGGGCGCAACACTTGGACA 9
XX
XX RESULT 11
XX AAX61766/c
XX ID AAX61766 standard; DNA; 2019 BP.
XX AC
XX AC AAX61766;
XX XX
XX XX 19-JUL-1999 (first entry)
XX DT
XX DE B. burgdorferi antigenic protein coding sequence, t679.nt.
XX KW
XX KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX OS
XX OS Borrelia burgdorferi.
XX XX
XX XX WO9859071-A1.
XX XX
XX XX 30-DEC-1998.
XX PD
XX PD 18-JUN-1998; 98WO-US12718.
XX PF
XX PF 03-SEP-1997; 97US-0057483.
XX PR
XX PR 20-JUN-1997; 97US-0050359.
XX PR
XX PR 22-JUL-1997; 97US-0053344.
XX PR
XX PR 22-JUL-1997; 97US-0053377.
XX XX
XX XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX PA (MEDI-) MEDIMUNE INC.
XX XX
XX XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX XX
XX XX WPI; 1999-189980/16.
XX DR
XX DR P-PSDB; AAY20069.
XX XX

```

PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases  
caused by Borrelia, particularly Lyme disease  
PS  
PS Claim 1; Page 184; 275pp; English.

CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the Borrelia genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the Borrelia genus. The products can also  
CC be used for detection of members of the Borrelia genus.

XX Sequence 2019 BP; 728 A; 204 C; 347 G; 740 T; 0 other;

Query Match 42.7%; Score 22.2; DB 20; Length 2019;  
Best Local Similarity 64.7%; Pred. No. 41;  
Matches 33; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 CAAACTAGTCTCAAGGTCATGCTCTTTAGGCCCAAACTAGGTCGAAGTC 51  
DB 1582 CAACACTAGGTCACAGCCCTCTCTTTAGCCCAATATCTATCAAGATC 1532

RESULT 12  
AAK61765/c  
ID AAK61765 standard; DNA; 2154 BP.  
AC AAK61765;  
XX  
XX 19-JUL-1999 (first entry)  
XX  
XX B. burgdorferi antigenic protein coding sequence, f679.nt.  
DE Antigenic protein; vaccine; Lyme disease; infection; detection; ss.  
KW Borrelia burgdorferi.  
OS  
OS WO9859071-A1.  
PN  
XX 30-DEC-1998.  
XX  
XX 18-JUN-1998; 98WO-US12718.  
XX  
XX 03-SEP-1997; 97US-0057483.  
PR 20-JUN-1997; 97US-0050359.  
PR 22-JUL-1997; 97US-0053344.  
PR 22-JUL-1997; 97US-0053377.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (MEDI-) MEDIMUNE INC.  
XX  
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;  
XX  
XX WPI; 1999-189980/16.  
DR P-PSDB; AAY20068.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop  
XX products for the diagnosis, prevention and treatment of diseases  
XX caused by Borrelia, particularly Lyme disease

PS Claim 1; Page 183-184; 275pp; English.

CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the Borrelia genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the Borrelia genus. The products can also  
CC be used for detection of members of the Borrelia genus.

XX Sequence 2154 BP; 768 A; 216 C; 363 G; 807 T; 0 other;

Query Match 42.7%; Score 22.2; DB 20; Length 2154;  
Best Local Similarity 64.7%; Pred. No. 42;  
Matches 33; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 CAAACTAGTCTCAAGGTCATGCTCTTTAGGCCCAAACTAGGTCGAAGTC 51  
DB 1717 CAACACTAGGTCACAGCCCTCTCTTTAGCCCAATATCTATCAAGATC 1667

RESULT 13  
AAQ02831/c  
ID AAQ02831 standard; DNA; 2716 BP.  
XX  
XX AAQ02831;

XX 29-MAY-1990 (first entry)  
XX  
XX Sequence encoding tumour necrosis factor.

XX Tumour necrosis factor.

XX Homo sapiens.

XX  
XX Key Location/Qualifiers  
XX FT sig\_peptide 214..297 /\*tag= a  
XX FT mat\_peptide 298..1464 /\*tag= b  
XX FT polyA\_site 2711 /\*tag= c

XX JP01285191-A.  
XX 16-NOV-1989.  
XX  
XX 13-MAY-1988; 88JP-0114921.  
XX  
XX 13-MAY-1988; 88JP-0114921.  
XX  
XX (GREC ) GREEN CROSS CORP.  
XX  
XX WPI; 1990-003177/01.  
XX P-PSDB; AAR01940.

XX DNA sequence, expression vector and host - of tumour necrosis factor.  
XX Disclosure; Fig 1; 11pp; Japanese.

XX The DNA was sequenced from cloned mRNA isolated from foetal kidney  
XX fibroblasts. The DNA can be integrated into an expression vector for  
XX transformation of CHO cells in order to produce tumour necrosis factor.

XX Sequence 2716 BP; 572 A; 721 C; 780 G; 643 T; 0 other;

Query Match 42.7%; Score 22.2; DB 11; Length 2716;  
Best Local Similarity 88.9%; Pred. No. 44;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 AGGTCAAAGGTCATGCTCTTTAGGCCCA 34  
DB 413 AGGTCAAAGGTCATGCTCTTTAGGCCCA 387

RESULT 14  
AAK20250/c  
ID AAK20250 standard; DNA; 111309 BP.  
XX  
XX AAK20250;

XX 04-MAY-1999 (first entry)

XX Borrelia burgdorferi polynucleotide sequence #3.

XX Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.  
 OS Borrelia burgdorferi.  
 XX WO9858943-A1.  
 XX 30-DEC-1998.  
 XX 18-JUN-1999; 98WO-US-2764.  
 XX 03-SEP-1997; 97US-0057483.  
 XX 20-JUN-1997; 97US-0050359.  
 XX 22-JUL-1997; 97US-0053344.  
 XX 22-JUL-1997; 97US-0053377.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 XX Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;  
 PI White OR;  
 XX WPI; 1999-081217/07.  
 XX New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease  
 XX Claim 1; Page 738-800; 11289p; English.  
 XX AAX20248 to AAX20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.  
 XX Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T; 10 other;  
 SQ  
 Query Match 42.7%; Score 22.2; DB 20; Length 111309;  
 Best Local Similarity 64.7%; Pred. No. 1.1e+02;  
 Matches 33; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
 QY 1 CAAACTAGGTCAAAGTCATGCTTTAGGCCCAAACTAGGTCAAAGGTC 51  
 |||||  
 Db 50401 CAACACTAGTTCAGCCCTCTCTTTAGCCCAATATTCATCAAGATC 50351  
 |||||  
 RESULT 15  
 AAC39133  
 ID AAC39133 standard; DNA; 1168 BP.  
 XX AC  
 XX AAC39133;  
 XX 17-OCT-2000 (first entry)  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 23495.  
 XX Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 XX EP1033405-A2.  
 XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.  
 XX 99US-0121825.  
 PR 25-FEB-1999; 99US-0123180.  
 PR 05-MAR-1999; 99US-0123548.  
 PR 09-MAR-1999; 99US-0125788.  
 PR 23-MAR-1999; 99US-0126364.  
 PR 25-MAR-1999; 99US-0126785.  
 PR 29-MAR-1999; 99US-0127462.  
 PR 01-APR-1999; 99US-0128234.  
 PR 06-APR-1999; 99US-0128714.  
 PR 08-APR-1999; 99US-0129845.  
 PR 16-APR-1999; 99US-0130077.  
 PR 19-APR-1999; 99US-0130449.  
 PR 21-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 30-APR-1999; 99US-0132484.  
 PR 04-MAY-1999; 99US-0132485.  
 PR 05-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135244.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139452.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139753.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140931.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0143920.  
PR 12-JUL-1999; 99US-0143977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145216.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147933.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159504.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161484.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 42.3%; Score 22; DB 21; Length 1168;  
Best Local Similarity 73.7%; Pred. No. 43;  
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 9 GGTCAAAGTCATGCTTTAGGCCCAAACTAGGTCAA 46  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1043 GTTCAAAGTCCTTTGTATAGGCCCTTAGGTAGGACAA 1080

Search completed: February 19, 2003, 21:23:05  
Job time : 53.6042 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: February 19, 2003, 20:28:04 ; Search time 6.54547 Seconds  
(without alignments)  
2399.710 Million cell updates/sec

file: US-09-808-388-4

effect score: 52

equences: 1 caaaactaggctcaaggtca.....caaaactaggctcaaggtca 52

coring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

searched: 441362 seqs, 153338381 residues

total number of hits satisfying chosen parameters: 882724

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgm2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgm2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgm2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgm2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgm2\_6/ptodata/2/ina/PTCUS\_COMB.seq.\*

6: /cgm2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
C 1	23.2	44.6	232	5	PCT-US93-06251-44
C 2	21.8	41.9	72928	3	US-09-009-913-1
C 3	20.8	40.0	541	4	US-03-404-879A-26
C 4	20.8	40.0	3970	1	US-07-925-695-3
C 5	20.8	40.0	9589	1	US-07-925-695-1
C 6	20.8	40.0	9589	1	US-07-925-695-2
C 7	20.6	39.6	2455	1	US-08-073-807A-1
C 8	20.6	39.6	4052	1	US-08-057-167-1
C 9	20.6	39.6	4052	5	PCT-US93-05412-1
C 10	20.6	39.6	245240	2	US-08-724-394A-20
C 11	20.6	39.6	245240	2	US-08-724-394A-21
C 12	20.6	39.6	245240	2	US-08-724-394A-22
C 13	20.4	39.2	3022	4	US-09-232-278A-8
C 14	20.4	39.2	3177	3	US-09-058-489-50
C 15	20.4	39.2	3674	1	US-08-105-483-324
C 16	20.4	39.2	3674	1	US-08-709-209-324
C 17	20.4	39.2	3674	1	US-08-458-101-324
C 18	20.2	38.8	326	1	US-08-458-550-5
C 19	20	38.5	4808	1	US-08-331-143-17
C 20	20	38.5	4808	2	US-08-025-563-17
C 21	20	38.5	4874	4	US-08-894-344C-1
C 22	20	38.5	9621	4	US-09-125-028-1
C 23	20	38.5	45546	4	US-09-146-053-6
C 24	19.8	38.1	1488	2	US-08-834-033A-3
C 25	19.8	38.1	1488	3	US-08-834-033A-3
C 26	19.8	38.1	1488	3	US-08-363-574-3
C 27	19.8	38.1	1488	4	US-08-363-526-3

28	19.8	38.1	1488	4	US-09-330-235-19	Sequence 19, Appl
29	19.8	38.1	1733	3	US-09-147-522-1	Sequence 1, Appl
30	19.8	38.1	2417	1	US-08-011-388B-1	Sequence 1, Appl
31	19.8	38.1	2417	1	US-08-464-051-1	Sequence 1, Appl
32	19.8	38.1	2417	2	US-08-462-498-1	Sequence 1, Appl
33	19.8	38.1	2417	3	US-08-554-385-2	Sequence 2, Appl
34	19.8	38.1	2886	4	US-09-280-116-104	Sequence 104, App
35	19.8	38.1	3511	2	US-08-727-118-1	Sequence 1, Appl
36	19.6	37.7	827	4	US-08-858-207A-190	Sequence 190, App
37	19.6	37.7	1497	1	US-08-488-961-5	Sequence 5, Appl
38	19.6	37.7	1497	4	US-08-973-297-5	Sequence 5, Appl
39	19.6	37.7	1497	5	PCT-US96-06511-5	Sequence 5, Appl
40	19.6	37.7	28882	4	US-08-961-527-140	Sequence 140, App
41	19.4	37.3	1846	1	US-08-483-389-117	Sequence 117, App
42	19.4	37.3	2986	3	US-09-042-416-1	Sequence 1, Appl
43	19.4	37.3	3016	2	US-08-344-155C-97	Sequence 97, Appl
44	19.4	37.3	3017	4	US-09-003-490A-86	Sequence 86, Appl
45	19.4	37.3	3024	6	5284931-1	Patent No. 5284931

## ALIGNMENTS

RESULT 1  
PCT-US93-06251-44/c  
; Sequence 44, Application PC/TUS9306251  
; GENERAL INFORMATION:  
; APPLICANT: Wickstrom, Eric and Rife, Jason P.  
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing Stereospecific Alkylphosphonates and Arylphosphonates  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
; STREET: 400 Garden City Plaza  
; CITY: Garden City  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 11530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/06251  
; FILING DATE: 19930630  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Digiglio, Frank S.  
; REGISTRATION NUMBER: 31,346  
; REFERENCE/DOCKET NUMBER: 8586  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 516-742-4343  
; TELEFAX: 516-742-4366  
; TELEX: 230 901 SANS UR  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 232 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
PCT-US93-06251-44

Query Match 44.6%; Score 23.2; DB 5; Length 232;  
Best Local Similarity 70.5%; Fred. No. 1.1;  
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
Qv 2 AAAAAGTAGGTCAAGGTGATGCTTTAGGCCCAAACTAGGTCA 45  
Db 73 AAAAAGTAGGTCTTTGGTGGGTTCTTTAGGCCCAAGCACTGGGTCA 30

## RESULT 2

US-09-009-913-1  
; Sequence 1, Application US/09009913  
; Patent No. 6087485  
; GENERAL INFORMATION:  
; APPLICANT: AxyS Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Asthma Related Genes  
; NUMBER OF SEQUENCES: 339  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bozicevic & Reed, LLP  
; STREET: 285 Hamilton Ave, Suite 200  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/009,913  
; FILING DATE: 21-JAN-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sherwood, Pamela J  
; REGISTRATION NUMBER: 36,677  
; REFERENCE/DOCKET NUMBER: SEQ-4P  
; TELEPHONE: 650-327-3231  
; TELEFAX: 650-327-3231  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 72928 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
US-09-009-913-1

Query Match 41.9%; Score 21.8; DB 3; Length 72928;  
Best Local Similarity 65.3%; Pred. No. 16; Indels 0; Gaps 0;  
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 AAAACTAGGTCAAAGGTCAATGCTTTTAGGCCCAAACTAGGTCAAAGGT 50  
DB 37573 AAAGCTGATTCATGTCATCTCTTTCTCAGCAATCTTGTCATGTGT 37621

## RESULT 3

US-09-404-879A-26/c  
; Sequence 26, Application US/09404879A  
; Patent No. 6468546  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Aigate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.462C2  
; CURRENT APPLICATION NUMBER: US/09/404,879A  
; NUMBER OF SEQ ID NOS: 393  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 26  
; LENGTH: 541  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-404-879A-26

Query Match 40.0%; Score 20.8; DB 4; Length 541;  
Best Local Similarity 78.1%; Pred. No. 12; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 6 CTAGGTCAAAGGTCAATGCTTTTAGGCCCAAA 37  
DB 294 CAAGGTCAAGGACATGCTTTTAGGCCCAAA 263

## RESULT 4

US-07-925-695-3  
; Sequence 3, Application US/07925695  
; Patent No. 5428145  
; GENERAL INFORMATION:  
; APPLICANT: OKAMOTO, Hiroaki  
; APPLICANT: KAKAMURA, Tetsuo  
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME,  
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND  
; TITLE OF INVENTION: DETECTION SYSTEMS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Beveridge, Degrandi, Weilacher & Young  
; STREET: 1850' M Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/925,695  
; FILING DATE: 19920807  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 287402/91  
; FILING DATE: 09-AUG-1991  
; PRIOR APPLICATION DATA: JP 360441/91  
; APPLICATION NUMBER: 05-DEC-1991  
; FILING DATE: 05-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weilacher, Robert G.  
; REGISTRATION NUMBER: 20,531  
; REFERENCE/DOCKET NUMBER: 06/87-48009  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 659-2811  
; TELEFAX: (202) 659-1462  
; TELEX: WUI 64470  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3970 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-07-925-695-3

Query Match 40.0%; Score 20.8; DB 1; Length 3970;  
Best Local Similarity 64.6%; Pred. No. 20; Indels 0; Gaps 0;  
Matches 31; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 5 ACTAGGTCAAAGGTCAATGCTTTTAGGCCCAAACTAGGTCAAAGGTCA 52  
DB 1954 ACTTGTCCAGGCACATGGCATCAATCCCAACATTAGGACTGGGTCA 2001

## RESULT 5

US-07-925-695-1  
; Sequence 1, Application US/07925695  
; Patent No. 5428145  
; GENERAL INFORMATION:





SEQUENCE CHARACTERISTICS:  
LENGTH: 2455 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 191..1438  
US-08-073-807A-1

Query Match 39.6%; Score 20.6; DB 1; Length 2455;  
Best Local Similarity 85.2%; Pred. No. 21;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 8 AGGTCAAGGTCATGCTTTAGGCCCA 34  
Db 387 AGGTCAAGGTCATGCTTTAGGCCCA 361

RESULT 8  
US-08-057-167-1/c  
Sequence 1, Application US/08057167  
Patent No. 5541095  
GENERAL INFORMATION:  
APPLICANT: Hirschberg, Carlos B.  
APPLICANT: Orellana, Ariel  
APPLICANT: Hashimoto, Yasuhiro  
APPLICANT: Swiedler, Stuart J.  
APPLICANT: Wei, Zheng  
APPLICANT: Ishihara, Masayuki  
TITLE OF INVENTION: GLYCOSAMINOGLYCAN SPECIFIC  
TITLE OF INVENTION: SULFOTRANSFERASES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 50Z or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/057,167  
FILING DATE: 19930430  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/899,432  
FILING DATE: 16 June 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: CLARK, PAUL T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 04020/015002  
TELEPHONE: (617) 542-8906  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4052  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-057-167-1

Query Match 39.6%; Score 20.6; DB 1; Length 4052;  
Best Local Similarity 62.7%; Pred. No. 24;  
Matches 32; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 1 CAAAAGTCAAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGGTC 51

Db 3836 CAGGACAGGCGCCCAAGGCCCATGCTCTGGGCCCATGGCAGGCGCAGTGCC 3786

RESULT 9  
PCT-US93-05412-1/c  
Sequence 1, Application PC/TUS9305412  
GENERAL INFORMATION:  
APPLICANT: Hirschberg, Carlos B.  
APPLICANT: Orellana, Ariel  
APPLICANT: Hashimoto, Yasuhiro  
APPLICANT: Swiedler, Stuart J.  
APPLICANT: Wei, Zheng  
APPLICANT: Ishihara, Masayuki  
TITLE OF INVENTION: GLYCOSAMINOGLYCAN SPECIFIC  
TITLE OF INVENTION: SULFOTRANSFERASES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 50Z or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/05412  
FILING DATE: 19930607  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/899,432  
FILING DATE: 16 June 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: CLARK, PAUL T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 04020/015002  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4052  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US93-05412-1

Query Match 39.6%; Score 20.6; DB 5; Length 4052;  
Best Local Similarity 62.7%; Pred. No. 24;  
Matches 32; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 1 CAAAAGTCAAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGGTC 51  
Db 3836 CAGGACAGGCGCCCAAGGCCCATGCTCTGGGCCCATGGCAGGCGCAGTGCC 3786

RESULT 10  
US-08-724-394A-20  
Sequence 20, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Krommal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchinashi, Zenta  
APPLICANT: Wolff, Roger K.

```
/ TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
/
/ NUMBER OF SEQUENCES: 31
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
/ STREET: Two Embarcadero Center, 8th Floor
/ CITY: San Francisco
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94111-3834
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/724,394A
/ FILING DATE: 01-OCT-1996
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fitts, Renee A.
/ REGISTRATION NUMBER: 35,136
/ REFERENCE/DOCKET NUMBER: 017957-000100
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-576-0300
/ TELEFAX: 415-576-0300
/ INFORMATION FOR SEQ ID NO: 20:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 246240 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: not relevant
/ TOPOLOGY: not relevant
/ MOLECULE TYPE: cdna
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 1..246240
/ OTHER INFORMATION: /note= "HLA-H. CONTIG"
/
/ US-08-724-394A-20
/
/ Query Match 39.6%; Score 20.6; DB 2; Length 246240;
/ Best Local Similarity 67.4%; Pred. No. 65;
/ Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
/
/ 20 1 CAAACTAGGTCAGAGGTCATGCTTTAGGCCCAAACTAGGT 43
/ Db 121059 CAAATAGGTGAGATTTTGTTTTAAGGATAAACTAGGT 121101
/
/ RESULT 11
/ US-08-724-394A-21
/ Sequence 21, Application US/08724394A
/ Patent No. 5872237
/ GENERAL INFORMATION:
/ APPLICANT: Feder, John N.
/ APPLICANT: Kronmal, Gregory S.
/ APPLICANT: Lauer, Peter M.
/ APPLICANT: Ruddy, David A.
/ APPLICANT: Thomas, Winston
/ APPLICANT: Tsuchihashi, Zenta
/ APPLICANT: Wolff, Roger K.
/ TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
/ TITLE OF INVENTION: Sequences and Antibodies Thereto
/ NUMBER OF SEQUENCES: 31
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
/ STREET: Two Embarcadero Center, 8th Floor
/ CITY: San Francisco
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94111-3834
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/724,394A
/ FILING DATE: 01-OCT-1996
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fitts, Renee A.
/ REGISTRATION NUMBER: 35,136
/ REFERENCE/DOCKET NUMBER: 017957-000100
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-576-0300
/ TELEFAX: 415-576-0300
```

```
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/724,394A
/ FILING DATE: 01-OCT-1996
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fitts, Renee A.
/ REGISTRATION NUMBER: 35,136
/ REFERENCE/DOCKET NUMBER: 017957-000100
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-576-0300
/ TELEFAX: 415-576-0300
/ INFORMATION FOR SEQ ID NO: 21:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 246240 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: not relevant
/ TOPOLOGY: not relevant
/ MOLECULE TYPE: cdna
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 1..246240
/ OTHER INFORMATION: /note= "HLA-H. CONTIG"
/
/ US-08-724-394A-21
/
/ Query Match 39.6%; Score 20.6; DB 2; Length 246240;
/ Best Local Similarity 67.4%; Pred. No. 65;
/ Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
/
/ 00 1 CAAACTAGGTCAGAGGTCATGCTTTAGGCCCAAACTAGGT 43
/ Db 121059 CAAATAGGTGAGATTTTGTTTTAAGGATAAACTAGGT 121101
/
/ RESULT 12
/ US-08-724-394A-22
/ Sequence 22, Application US/08724394A
/ Patent No. 5872237
/ GENERAL INFORMATION:
/ APPLICANT: Feder, John N.
/ APPLICANT: Kronmal, Gregory S.
/ APPLICANT: Lauer, Peter M.
/ APPLICANT: Ruddy, David A.
/ APPLICANT: Thomas, Winston
/ APPLICANT: Tsuchihashi, Zenta
/ APPLICANT: Wolff, Roger K.
/ TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
/ TITLE OF INVENTION: Sequences and Antibodies Thereto
/ NUMBER OF SEQUENCES: 31
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
/ STREET: Two Embarcadero Center, 8th Floor
/ CITY: San Francisco
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94111-3834
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/724,394A
/ FILING DATE: 01-OCT-1996
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Fitts, Renee A.
/ REGISTRATION NUMBER: 35,136
/ REFERENCE/DOCKET NUMBER: 017957-000100
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-576-0300
/ TELEFAX: 415-576-0300
```

INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cdna  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H. CONTIG"

US-08-724-394A-22

Query Match 39.6%; Score 20.6; DB 2; Length 246240;  
Best Local Similarity 67.4%; Pred. No. 65;  
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1 CAAGACTAGGTCAGGTCATGCTTTAGGCCCAAACTAGGT 43  
Db 121059 CAAGACTAGGTCAGGTCATGCTTTAGGCCCAAACTAGGT 121101

RESULT 13

US-09-232-278A-8  
Sequence 8, Application US/09232278A  
Patent No. 6348196

GENERAL INFORMATION:  
APPLICANT: AUDONNET et al.  
TITLE OF INVENTION: FELINE POLYNUCLEOTIDE VACCINE FORMULA  
FILE REFERENCE: 454313-2220  
CURRENT APPLICATION NUMBER: US/09/232,278A  
CURRENT FILING DATE: 1999-01-15  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 8  
LENGTH: 3022  
TYPE: DNA  
ORGANISM: Feline leukemia virus  
US-09-232-278A-8

Query Match 39.2%; Score 20.4; DB 4; Length 3022;  
Best Local Similarity 65.2%; Pred. No. 27;  
Matches 30; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 4 AACTAGGTCAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGG 49  
Db 1010 AAGAGGCAAGGTCCTTTTGGGCCCAAGAGGATCCAGG 1055

RESULT 14

US-09-058-489-50/c  
Sequence 50, Application US/09058489  
Patent No. 6103886

GENERAL INFORMATION:  
APPLICANT: Whitehead Institute for Biomedical Research  
APPLICANT: Lahn, Bruce  
APPLICANT: Page, David  
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of the Y Chromosome  
FILE REFERENCE: WHI97-08DA  
CURRENT APPLICATION NUMBER: US/09/058,489  
CURRENT FILING DATE: 1998-04-10  
EARLIER APPLICATION NUMBER: 60/041,877  
EARLIER FILING DATE: 1997-04-11  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 50  
LENGTH: 3177  
TYPE: DNA  
ORGANISM: Human  
US-09-058-489-50

Query Match 39.2%; Score 20.4; DB 3; Length 3177;

Best Local Similarity 71.1%; Pred. No. 27;  
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 8 AGGTCAAAGGTCATGCTTTAGGCCCAAACTAGGTCA 45  
Db 600 AGTCAAGCTCAGGTCATGCTTTAGGCCCAAGCTAGCACA 563

RESULT 15

US-08-105-483-324  
Sequence 324, Application US/08105483  
Patent No. 5494807

GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE  
TITLE OF INVENTION: STRAIN  
NUMBER OF SEQUENCES: 462  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
ADDRESSEE: c/o William S. Frommer  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/105,483  
FILING DATE: 12-AUG-1993  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/847,951  
FILING DATE: 06-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2400  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 324:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3674 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-105-483-324

Query Match 39.2%; Score 20.4; DB 1; Length 3674;  
Best Local Similarity 65.2%; Pred. No. 28;  
Matches 30; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 4 AACTAGGTCAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGG 49  
Db 1661 AAGAAAGCAAGGTCCTTTTGGGCCCAAGAGGATCCAGG 1706

Search completed: February 20, 2003, 01:16:19  
Job time : 160.645 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

DN nucleic - nucleic search, using sw model

Run on: February 19, 2003, 23:17:14 ; Search time 39.444 Seconds  
(without alignments)  
671.453 Million cell updates/sec

Title: US-09-808-388-4  
Perfect score: 52  
Sequence: 1 CAAACTAGGTCAAAGGTCA.....CAAACTAGGTCAAAGGTCA 52

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	52	10	US-09-808-388-4
2	23	44.2	867	10	US-09-770-445-571
3	22.8	43.8	60	9	US-09-877-705A-142
4	22.8	43.8	60	9	US-09-877-738A-142
5	22.2	42.7	42999	9	US-09-799-462A-17
6	22.2	42.7	42999	9	US-10-125-767-17
7	21.8	41.9	352	10	US-09-770-791-681
8	21.8	41.9	2086	10	US-09-764-877-3431
9	21.8	41.9	6027	10	US-09-070-927A-124
10	21.4	41.2	2480	9	US-10-078-770-53
11	21.4	41.2	2982	9	US-09-938-842A-2525
12	21.2	40.8	924	10	US-09-815-242-4203
13	21.2	40.8	960	10	US-09-815-242-8214
14	21.2	40.8	1008	10	US-09-876-225-1
15	21.2	40.8	73308	10	US-09-954-456-2276
16	21	40.4	41	10	US-09-808-388-3
17	21	40.4	332	10	US-09-808-388-6
18	21	40.4	478	10	US-09-864-761-580
19	21	40.4	508	10	US-09-783-590-4401

20 20.8 40.0 498 9 US-10-202-193-218 Sequence 218, Appl  
21 20.8 40.0 541 10 US-09-884-441-26 Sequence 26, Appl  
22 20.8 40.0 1312 9 US-10-202-193-335 Sequence 335, Appl  
23 20.8 40.0 2497 9 US-09-989-920-60 Sequence 60, Appl  
24 20.6 39.6 509 10 US-09-815-343-1278 Sequence 1278, Ap  
25 20.6 39.6 998 10 US-09-765-213A-1 Sequence 1, Appli  
26 20.6 39.6 1773 9 US-09-938-842A-2960 Sequence 2960, Ap  
27 20.6 39.6 2072 10 US-09-725-735A-11 Sequence 11, Appl  
28 20.6 39.6 4436 9 US-09-522-334-26 Sequence 26, Appl  
29 20.6 39.6 5046 10 US-09-725-735A-13 Sequence 13, Appl  
30 20.6 39.6 180557 12 US-10-003-806-6 Sequence 6, Appli  
31 20.6 39.6 180557 12 US-10-003-806-6 Sequence 6, Appli  
32 20.4 39.2 179 10 US-09-924-035A-658 Sequence 658, App  
33 20.4 39.2 433 10 US-09-783-590-7042 Sequence 7042, Ap  
34 20.4 39.2 776 10 US-09-910-943-407 Sequence 407, App  
35 20.4 39.2 1200 10 US-09-887-576-752 Sequence 752, App  
36 20.4 39.2 2000 9 US-09-938-842A-2897 Sequence 2897, Ap  
37 20.2 38.8 38 10 US-09-808-388-2 Sequence 2, Appli  
38 20.2 38.8 543 10 US-09-925-297-397 Sequence 397, App  
39 20.2 38.8 926 10 US-09-974-300-4809 Sequence 4809, Ap  
40 20.2 38.8 936 10 US-09-815-242-6754 Sequence 6754, Ap  
41 20.2 38.8 1201 10 US-09-205-448-3 Sequence 3, Appli  
42 20.2 38.8 2205 12 US-10-036-328A-3 Sequence 3, Appli  
43 20.2 38.8 3214 10 US-09-070-927A-280 Sequence 280, App  
44 20.2 38.8 3633 12 US-10-036-328A-1 Sequence 1, Appli  
45 20.2 38.8 4185 12 US-10-036-328A-9 Sequence 9, Appli

## ALIGNMENTS

RESULT 1  
US-09-808-388-4  
; Sequence 4, Application US/09808388  
; Patent No. US20020081719A1  
; GENERAL INFORMATION:  
; APPLICANT: Massaad, Charbel  
; APPLICANT: Berenbaum, Francis  
; APPLICANT: Olivier, Jean-Luc  
; APPLICANT: Salvat, Colette  
; APPLICANT: Berezat, Gilbert  
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
; TITLE OF INVENTION: their uses  
; FILE REFERENCE: ST00010  
; CURRENT APPLICATION NUMBER: US/09/808,388  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: FR/00/03262  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: US 60/196,959  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 52  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: PPRE element  
; US-09-808-388-4

Query Match 100.0%; Score 52; DB 10; Length 52;  
Best Local Similarity 100.0%; Pred. No. 6.3e-11;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAAAGGTCAATGCTTTAGGCCCAAACTAGGTCAAAGGTCA 52  
|||||  
Db 1 CAAACTAGGTCAAAGGTCAATGCTTTAGGCCCAAACTAGGTCAAAGGTCA 52  
|||||

RESULT 2  
US-09-770-445-571  
; Sequence 571, Application US/09770445  
; Patent No. US20020023281A1

GENERAL INFORMATION:  
APPLICANT: Gorlach, Jorn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Yang  
APPLICANT: Rameaka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Woessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Kricker, Maya  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hurban, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
TITLE OF INVENTION: thaliana  
FILE REFERENCE: 2023US (PARA-012PRV)  
CURRENT APPLICATION NUMBER: US/09/770,445  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: US 60/178,472  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 571  
LENGTH: 867  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(867)  
OTHER INFORMATION: n = A,T,C or G  
US-09-770-445-571

Query Match 44.2%; Score 23; DB 10; Length 867;  
Best Local Similarity 70.7%; Pred. No. 10;  
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 7 TAGGTCGAAGGTCATGCTTTAGGCCCAAACTAGGTCAAA 47  
Db 606 TAGGTTACCAATCAGCTCTTTGGCGCAACTAGATNAA 646

RESULT 3  
US-09-877-705A-142/c  
Sequence 142, Application US/09877705A  
Publication No. US20030008283A1  
GENERAL INFORMATION:  
APPLICANT: Li, Jason  
TITLE OF INVENTION: METHOD FOR SCREENING FOR DRUG CANDIDATES FOR MODULATING TRANSCRIPTION ACTIVITY  
TITLE OF INVENTION: FACTOR ACTIVITY  
FILE REFERENCE: 26757-704  
CURRENT APPLICATION NUMBER: US/09/877,705A  
CURRENT FILING DATE: 2001-08-16  
NUMBER OF SEQ ID NOS: 162  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 142  
LENGTH: 60  
TYPE: DNA  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Hybridization probe MP68  
US-09-877-705A-142

Query Match 43.8%; Score 22.8; DB 9; Length 60;  
Best Local Similarity 75.0%; Pred. No. 6.4;  
Matches 45; Conservative 0; Mismatches 8; Indels 1;

Qy 1 CAAACTAGGTCAAAGGTC-----ATGCTTTAGGCCCAAACTAGGTCAAAGGTCA 52  
Db 60 CAAACTAGGTCAAAGGTCACAAAAGGTCACAAAAGGTCACAAAAGGTCACAAAAGGTCA 1

RESULT 4  
US-09-877-738A-142/c  
Sequence 142, Application US/09877738A  
Publication No. US20030022173A1  
GENERAL INFORMATION:  
APPLICANT: Li, Jason  
TITLE OF INVENTION: METHOD AND KIT FOR ISOLATING DNA PROBES THAT BIND TO ACTIVATED TRANSCRIPTION FACTORS  
TITLE OF INVENTION: TRANSCRIPTION FACTORS  
FILE REFERENCE: 26757-701  
CURRENT APPLICATION NUMBER: US/09/877,738A  
CURRENT FILING DATE: 2001-06-01  
NUMBER OF SEQ ID NOS: 162  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 142  
LENGTH: 60  
TYPE: DNA  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Hybridization probe MP68  
US-09-877-738A-142

Query Match 43.8%; Score 22.8; DB 9; Length 60;  
Best Local Similarity 75.0%; Pred. No. 6.4;  
Matches 45; Conservative 0; Mismatches 7; Indels 8; Gaps 1;

Qy 1 CAAACTAGGTCAAAGGTC-----ATGCTTTAGGCCCAAACTAGGTCAAAGGTCA 52  
Db 60 CAAACTAGGTCAAAGGTCACAAAAGGTCACAAAAGGTCACAAAAGGTCACAAAAGGTCA 1

RESULT 5  
US-09-799-462A-17  
Sequence 17, Application US/09799462A  
Patent No. US20020160970A1  
GENERAL INFORMATION:  
APPLICANT: Hadlaczky, Gyula  
Szalay, Aladar  
TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/799,462A  
FILING DATE: 10-Sep-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/835,682  
FILING DATE: 10-APR-1997  
APPLICATION NUMBER: 08/695,191  
FILING DATE: 07-AUG-1996  
APPLICATION NUMBER: 08/682,080  
FILING DATE: 15-JUL-1996  
APPLICATION NUMBER: 08/629,822  
FILING DATE: 10-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 24601-402G  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858-450-8403  
TELEFAX: 858-587-5360  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4299 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: <Unknown>  
ORIGINAL SOURCE:  
SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-09-799-462A-17

Query Match 42.7%; Score 22.2; DB 9; Length 42999;  
Best Local Similarity 69.8%; Pred. No. 51;  
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

3 AAACAGTCAAGGTCTCTTTAGGCCCAAACTAGTCA 45  
35024 AAACAGTCAAGGTCTCTTTAGGCCCAAACTAGTCA 35066

RESULT 6  
US-10-125-767-17  
Sequence 17, Application US/10125767  
Patent No. US20020160410A1  
GENERAL INFORMATION:  
APPLICANT: Hadlaczky, Gyula  
TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS  
FOR PREPARING ARTIFICIAL CHROMOSOMES  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Heller Ehrman White & McAuliffe LLP  
STREET: 4350 La Jolla Village Drive, 7th Floor  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10125.767  
FILING DATE: 17-Apr-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/724,693  
FILING DATE: 28-NOV-2000  
APPLICATION NUMBER: 08/835,682  
FILING DATE: 10-APR-1997  
APPLICATION NUMBER: 08/695,191  
FILING DATE: 07-AUG-1996  
APPLICATION NUMBER: 08/682,080  
FILING DATE: 15-JUL-1996  
APPLICATION NUMBER: 08/629,822  
FILING DATE: 10-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 24601-402J  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858-450-8403  
TELEFAX: 858-587-5360  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4299 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: <Unknown>  
ORIGINAL SOURCE:  
SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-10-125-767-17

Query Match 42.7%; Score 22.2; DB 9; Length 42999;  
Best Local Similarity 69.8%; Pred. No. 51;  
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

3 AAACAGTCAAGGTCTCTTTAGGCCCAAACTAGTCA 45  
35024 AAACAGTCAAGGTCTCTTTAGGCCCAAACTAGTCA 35066

RESULT 7  
US-09-770-791-681  
Sequence 681, Application US/09770791  
Patent No. US20020062014A1  
GENERAL INFORMATION:  
APPLICANT: Goriach, Jorn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Yang  
APPLICANT: Rameaka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Woessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Krickler, Maja  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hurban, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2029 (PARA-018PRV)  
CURRENT APPLICATION NUMBER: US/09/770,791  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/178,480  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 681  
LENGTH: 352  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(352)  
OTHER INFORMATION: n = A,T,C or G  
US-09-770-791-681

Query Match 41.9%; Score 21.8; DB 10; Length 352;  
Best Local Similarity 65.3%; Pred. No. 23;  
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

2 AAACAGTCAAGGTCTCTTTAGGCCCAAACTAGTCAAGGT 50  
266 AAACAGTCAAGGTCTCTTTAGGCCCAAACTATCAATGT 314

```
RESULT 8
US-09-764-877-3431/c
; Sequence 3431, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3431
; LENGTH: 2086
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3431

Query Match 41.9%; Score 21.8; DB 10; Length 2086;
Best Local Similarity 70.7%; Pred. No. 35;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

CY 2 AACTAGGTCAGGTCATGCTCTTTAGGCCCAAACTAGG 42
Db 1821 AAAAAAGCCAAAGGAGATTCTTGTGGAAAAAACTAAG 1781

RESULT 9
US-09-070-927A-124
; Sequence 124, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; APPLICANT: Patrick J. Dillon
; APPLICANT: Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 124:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6027 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 124:
US-09-070-927A-124

Query Match 41.9%; Score 21.8; DB 10; Length 6027;
Best Local Similarity 70.7%; Pred. No. 45;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

CY 12 CAAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGGTCA 52
Db 5861 CAAGGTAATTTCTTGTCTCCCAATATATTGAAACGGTCA 5901

RESULT 10
US-10-078-770-53
; Sequence 53, Application US/10078770
; Publication No. US20030003471A1
; GENERAL INFORMATION:
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Forge, Charlie
; APPLICANT: Miao, Guo-Hua
; TITLE OF INVENTION: cDNAs Encoding Polypeptides
; FILE REFERENCE: BB-1365 US NA
; CURRENT APPLICATION NUMBER: US/10/078,770
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/614,188
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,400
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/153,534
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/161,223
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/159,878
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: 60/157,401
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 60/143,419
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,409
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 53
; LENGTH: 2480
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-078-770-53

Query Match 41.2%; Score 21.4; DB 9; Length 2480;
Best Local Similarity 66.0%; Pred. No. 52;
Matches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

CY 4 AACTAGTCAAAGGTCATGCTCTTTAGGCCCAAACTAGGTCAAAGGT 50
Db 1453 AACTAGTCAAAGGAATATTATTGATGTCACACTACCTCACAAGT 1499

RESULT 11
US-09-938-842A-2525/c
; Sequence 2525, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
```

PRIOR APPLICATION NUMBER: US 60/227,866  
 PRIOR FILING DATE: 2000-08-24  
 PRIOR APPLICATION NUMBER: US 60/264,647  
 PRIOR FILING DATE: 2001-01-16  
 PRIOR APPLICATION NUMBER: US 60/300,111  
 PRIOR FILING DATE: 2001-06-22  
 NUMBER OF SEQ ID NOS: 5379  
 SEQ ID NO 2525  
 LENGTH: 2982  
 TYPE: DNA  
 ORGANISM: Arabidopsis thaliana  
 S-09-938-842A-2525

Query Match 41.2%; Score 21.4; DB 9; Length 2982;  
 Best Local Similarity 66.0%; Pred. No. 54;  
 Matches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Y 3 AACTAGTCAAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGG 49  
 |||||  
 b 1133 AATGAGGCGAAACTCGGTATAGAGTCCCAAGACTGGGTCAAAGG 1087

# ESULT 12

S-09-815-242-4203/c  
 Sequence 4203, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 4203  
 LENGTH: 924  
 TYPE: DNA  
 ORGANISM: Staphylococcus aureus  
 S-09-815-242-4203

Query Match 40.8%; Score 21.2; DB 10; Length 924;  
 Best Local Similarity 64.0%; Pred. No. 49;  
 Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Y 1 CAAACTAGTCAAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGGT 50  
 |||||  
 b 84 CAAAATTTATCTAAAGCCATACCTATAGATCAAGCCAACTCCAGGT 35

# ESULT 13

S-09-815-242-8214/c

; Sequence 8214, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 8214  
 LENGTH: 960  
 TYPE: DNA  
 ORGANISM: Staphylococcus aureus

FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)...(960)

US-09-815-242-8214

Query Match 40.8%; Score 21.2; DB 10; Length 960;  
 Best Local Similarity 64.0%; Pred. No. 49;  
 Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 CAAACTAGTCAAGGTCATGCTTTAGGCCCAAACTAGGTCAAAGGT 50  
 |||||  
 Db 138 CAAAATTTATCTAAAGCCATACCTATAGATCAAGCCAACTCCAGGT 89

# RESULT 14

US-09-876-225-1  
 ; Sequence 1, Application US/09876225  
 ; Patent No. US20020052025A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baylor College of Medicine  
 ; APPLICANT: Thompson, Timothy C.

TITLE OF INVENTION: RTVP Based Compositions and Methods for the Treatment of Prostate

FILE REFERENCE: 38594.0027  
 CURRENT APPLICATION NUMBER: US/09/876,225  
 CURRENT FILING DATE: 2001-06-08  
 PRIOR APPLICATION NUMBER: US 60/209,989  
 PRIOR FILING DATE: 2000-06-08  
 NUMBER OF SEQ ID NOS: 5  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 1  
 LENGTH: 1008  
 TYPE: DNA  
 ORGANISM: Mus sp.

US-09-876-225-1

Query Match

40.8%; Score 21.2; DB 10; Length 1008;



```

Best Local Similarity 76.5%; Pred. No. 50;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 19 CATGCTTTAGGCCCAAACTAGGTCAAGGTCA 52
Db 322 CATGCTTTGGGACCCCAAACTAGGCCCAAACTGCA 355

RESULT 15
US-09-954-456-2276/c
; Sequence 2276, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2276
; LENGTH: 73308
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-2276

Query Match 40.8%; Score 21.2; DB 10; Length 73309;
Best Local Similarity 64.0%; Pred. No. 1.4e+02;
Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 3 AAACAGGTCAAGGTCAAGTCTTTAGGCCCAAACTAGGTCAAGGTCA 52
Db 56065 AAACAGGTCAAGGTCAAGTCTTTAGGCCCAAACTAGGTCAAGGTCA 56016

Search completed: February 20, 2003, 07:01:47
Job time : 66.444 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:33:48 ; Search time 253.018 Seconds  
(without alignments)  
3328.484 Million cell updates/sec

Title: US-09-808-388-4

Perfect score: 52

Sequence: 1 caaaactaggtcaagggtca.....caaaactaggtcaagggtca 52

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_estba.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_estc.\*

9: gb\_estl.\*

10: gb\_estc.\*

11: gb\_estc.\*

12: gb\_estc.\*

13: gb\_estc.\*

14: gb\_estc.\*

15: em\_estfun.\*

16: em\_estom.\*

17: gb\_gss.\*

18: em\_gss\_hum.\*

19: em\_gss\_inv.\*

20: em\_gss\_pln.\*

21: em\_gss\_vrt.\*

22: em\_gss\_fun.\*

23: em\_gss\_mam.\*

24: em\_gss\_mus.\*

25: em\_gss\_other.\*

26: em\_gss\_pro.\*

27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	27	51.9	457	17	BH595543 BOHSA82R
C 2	27	51.9	588	17	AQ289893 nbxb0036A
C 3	26.4	50.8	1036	17	AL184317 Tetraodon
C 4	26.2	50.4	485	17	AZ154378 SP_0032 A
C 5	25.8	49.6	609	17	BH485024 BOGFL15TR
C 6	25.8	49.6	839	17	BH666577 BOMEH39TR

7	25.8	49.6	1042	17	BH694652
8	25.6	49.2	423	14	BM859128
9	25.4	48.8	1062	17	CNS04TSL
C 10	25.2	48.5	719	17	AZ315041
11	25	48.1	407	10	AW875156
12	24.6	47.3	606	9	AA870634
C 13	24	46.2	167	10	BE483332
C 14	24	46.2	285	10	BE483320
C 15	24	46.2	338	12	BF707175
C 16	24	46.2	338	12	BF707176
C 17	24	46.2	340	13	EM436148
18	24	46.2	372	10	AV666466
C 19	24	46.2	402	13	EM433145
C 20	24	46.2	426	10	AV589752
C 21	24	46.2	426	12	BG224317
C 22	24	46.2	427	12	BG223619
C 23	24	46.2	436	10	AM670137
C 24	24	46.2	438	10	BE478607
C 25	24	46.2	441	10	AV590502
C 26	24	46.2	458	10	BE482505
C 27	24	46.2	460	10	AM670144
C 28	24	46.2	469	10	BE588867
C 29	24	46.2	480	12	BG691126
C 30	24	46.2	504	12	BF046014
C 31	24	46.2	521	13	EM366351
C 32	24	46.2	533	10	AV666467
C 33	24	46.2	567	10	AV590032
C 34	23.8	45.8	444	13	BI345006
C 35	23.8	45.8	688	13	BI178869
C 36	23.8	45.8	726	13	BI179191
C 37	23.8	45.8	726	14	BO509752
C 38	23.8	45.8	804	17	BH047011
39	23.6	45.4	425	10	BE246237
40	23.6	45.4	474	17	DR11A18T
C 41	23.6	45.4	520	14	BQ768648
C 42	23.4	45.0	446	17	AQ110493
C 43	23.4	45.0	625	17	AZ640770
C 44	23.4	45.0	663	10	AV375316
C 45	23.4	45.0	670	13	BI956025

## ALIGNMENTS

RESULT 1  
BH595543/c  
LOCUS BOHSA82TR BOHS Brassica oleracea genomic clone BOHSA82, DNA  
DEFINITION BH595543  
sequence.  
ACCSSION BH595543  
VERSION BH595543.1 GI:17847995  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 457)  
TOWN,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
JOURNAL Other GSSs: BOHSA82TF  
COMMENT Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1. .457

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1. .457

/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHSA82"  
/clone\_lib="BOHS"  
/note="Vector: PHOS1, Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"  
BASE COUNT 140 a 99 c 74 g 144 t  
ORIGIN

Query Match 51.9%; Score 27; DB 17; Length 457;  
Best Local Similarity 70.6%; Pred. No. 6.7;  
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CAAACTAGGTCAGGTCATGCTTTAGGCCCAAACTAGGTCAGGTC 51  
Db 199 CATAGTTGTCAGAGGCTCTAGCTATAGATCATAGGTCAGGTC 149

RESULT 2  
AQ289893/c  
LOCUS  
DEFINITION  
nbx0036A07f: CUGI Rice BAC Library Oryza sativa genomic clone  
nbx0036A07f, DNA sequence.

ACCESSION  
AQ289893  
VERSION  
AQ289893.1 GI:3951255  
KEYWORDS  
GSS.  
SOURCE  
Oryza sativa.  
ORGANISM  
Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 588)  
Wing, R.A. and Dean, R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)  
Contact: Wing RA  
Clemson University  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATAGGACTCACTATAGG  
Class: BAC ends  
High quality sequence stop: 468.  
Location/Qualifiers  
1..588  
/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="nbx0036A07f"  
/clone\_lib="CUGI Rice BAC Library"  
/issue\_type="Leaf"  
/lab\_host="E. coli DH10B"  
/note="Vector: pBeloBAC11; Site 1: HindIII; Site 2:  
HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters,

each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."  
BASE COUNT 150 a 100 c 117 g 220 t  
ORIGIN

Query Match 51.9%; Score 27; DB 17; Length 588;  
Best Local Similarity 70.6%; Pred. No. 7.3;  
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 AAAACTAGGTCAGGTCATGCTTTAGGCCCAAACTAGGTCAGGTC 52  
Db 452 AGAGCTGGTCAACAGTCATTTTGTGAGGCCCAATGTAGATAACATGTC 402

RESULT 3  
CNS02770/c  
LOCUS  
DEFINITION  
CNS02770  
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
242314 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

ACCESSION  
AL184317  
VERSION  
AL184317.1 GI:7822421  
KEYWORDS  
GSS; genome survey sequence.  
SOURCE  
Tetraodon nigroviridis.  
ORGANISM  
Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 1036)  
Roest-Crolius, H., Jaillon, O., Dasilva, C., Bounau, L., Fisher, C.,  
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,  
Saurin, W. and Weissenbach, J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
2 (bases 1 to 1036)  
Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,  
Bounau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and  
Weissenbach, J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
3 (bases 1 to 1036)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000)  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/tetraodon.  
Location/Qualifiers  
1..1036  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="242314"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG242BD07SP1-end :  
PUC-Ori"

BASE COUNT 248 a 286 c 212 g 291 t  
ORIGIN

Query Match 50.8%; Score 26.4; DB 17; Length 1036;  
Best Local Similarity 75.0%; Pred. No. 15;  
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 9 GGTCAGGTCATGCTTTAGGCCCAAACTAGGTCAGGTC 52  
Db 615 GATCAAGGCTCCTTGTGCTTTGGCCCTAAATAGTTCAGGTC 572

RESULT 4  
AZ154378/c





freshwater pufferfish Tetraodon nigroviridis

JOURNAL  
REFERENCE 3 (bases 1 to 1062)  
Genoscope.

TITLE  
JOURNAL Direct Submission

COMMENT Submitted (12-APR-2000)

This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

Location/Qualifiers

FEATURES

source

1. 1062

/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone="038113"

/clone\_lib="A"

/notes="Genoscope sequence ID : C1AAC007ZE04A1-end : T3"

BASE COUNT 315 a 154 c 258 g 283 t 52 others

ORIGIN

Query Match 48.8%; Score 25.4; DB 17; Length 1062;

Best Local Similarity 72.7%; Pred. No. 36;

Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Cy 9 GGTCAAGTCTATCTTTAGGCCCAAACTAGTCAAGGTCA 52

Db 229 GATCAAGCTCTTTGGTTTGGCCCTAAANAGTCAAGGTCA 272

RESULT 10

AZ315041/c

LOCUS

AZ315041 719 bp DNA linear GSS 29-SEP-2000

DEFINITION IM0032B12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0032B12 F, DNA sequence.

ACCESSION

AZ315041

VERSION

AZ315041.1 GI:10361500

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 719)

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

JOURNAL

Unpublished (2000)

CONTACT: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Insert Length: 10000 Std Error: 0.00

Plate: 0032 row: B column: 12

Seq primer: CGTGTAAACGACGCGCCAGT

Class: plasmid ends

High quality sequence stop: 719.

Location/Qualifiers

1. 719

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0032B12"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

source

FEATURES

source

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (GI:4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and

purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells

and selected for ampicillin resistance." BASE COUNT 179 a 197 c 151 g 192 t

ORIGIN

Query Match 48.5%; Score 25.2; DB 17; Length 719;

Best Local Similarity 78.9%; Pred. No. 38;

Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Cy 15 AGGTCTGCTTTAGGCCCAAACTAGTCAAGGTCA 52

Db 718 AGAGCATGTCATTACACCAAACTAGGCAAGGTCA 681

RESULT 11

AW875156

LOCUS

AW875156 407 bp mRNA linear EST 22-MAY-2000

DEFINITION RCO-PT0006-271199-011-B08 PT0006 Homo sapiens cDNA, mRNA sequence.

ACCESSION

AW875156

VERSION

AW875156.1 GI:8012925

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 407)

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

20202663

MEDLINE

COMMENT

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=8t2=RCO-PT0006-271>)

199-011-B08&t3=1999-11-27&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 154.

Location/Qualifiers

1. 407

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="PT0006"

/dev\_stage="Adult"

/note="Organ: pnet; Vector: puc18; Site 1: SmaI; Site 2:

SmaI; A mini-library was made by cloning products derived

source

FEATURES

source

from ORSTES PCR (U.S. Letters Patent application No. 196  
716 - Ludwig Institute for Cancer Research) profiles  
into the pUC19 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 145 a 76 c 122 g 64 t  
ORIGIN

Query Match 48.1%; Score 25; DB 10; Length 407;

Best Local Similarity 69.4%; Pred. No. 37; Mismatches 0; Indels 0; Gaps 0;

Matches 34; Conservative 0;

QY 1 CAAGTCTAGTCAAGTCTAGTCTTTAGGCCCAAACTAGTCAAGG 49

Db 268 CAAAAATTGGAAAGGGCTTCCCTTGCCCAAAACAATGACTAAGG 316

# RESULT 12

AA870634

LOCUS 606 bp mRNA linear EST 16-MAR-1998  
DEFINITION vq24b07.r1 Barstead stromal cell line MPLRB8 Mus musculus cDNA  
clone IMAGE:1095157 5' similar to gb:X80899 M.musculus mRNA for  
cytochrome C oxidase subunit VII (MOUSE);, mRNA sequence.

ACCESSION AA870634

VERSION AA870634.1 GI:2966079

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 606)  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lucy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:601389

Seq primer: -28mi3 rev2 ET from Amersham

High quality sequence stop: 338.

Location/Qualifiers

FEATURES

source

1..606

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:1095157"

/clone\_lib="Barstead stromal cell line MPLRB8"

/cell\_line="C2C12 (undifferentiated)"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site1: EcoRI; Site2: NotI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TTTTCAGTACATGAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

[AATTCGGATCCTTG], digested with Not I and cloned into the

Not I and Eco RI sites of the modified pT7T3 vector.

Source undifferentiated tissue culture cell line C2C12.

Library constructed by Bob Barstead. The C2C12 cell line

(available from ATCC, catalog # CRL-1772) differentiates

rapidly, forming contractile myotubes and producing

characteristic muscle proteins. "

151 a 120 c 172 g 163 t

BASE COUNT

ORIGIN

Query Match 47.3%; Score 24.6; DB 9; Length 606;

Best Local Similarity 76.9%; Pred. No. 60;

Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 11 TCAAGGTCATGCTTTAGGCCCAAACTAGTCAAGG 49

Db 464 TCAAGGTTGTGTCTGCGCCACACCTAGGGGAAGG 502

# RESULT 13

BE483332/c

LOCUS

DEFINITION 169467 BARC 5BOV Bos taurus cDNA 5', mRNA linear EST 28-AUG-2000

ACCESSION BE483332

VERSION BE483332.1 GI:9602865

KEYWORDS EST.

SOURCE cow.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 167)

REFERENCE Sonstegard, T.S., Capuco, A.V., Van Tassel, C.P., Ashwell, M.S. and

Wells, K.D.

Mapping of Expressed Sequence Tags from a normalized bovine mammary

gland cDNA library

Unpublished (2000)

Contact: Sonstegard TS

USDA, ARS, Beltsville Agricultural Research Center

Bldg. 200 Rm 2A, Beltsville, MD 20705, USA

Tel: 301 504 8416

Fax: 301 504 8414

Email: tads@nri.barc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred

v0.980904.e. Vector identified by cross\_match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 13 row: P column: 3

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1..167

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="BARC 5BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;

Library made from pooled mRNA isolated from mammary

tissues at eight physiological, developmental, and disease

states."

BASE COUNT 46 a 41 c 36 g 44 t

ORIGIN

Query Match 46.2%; Score 24; DB 10; Length 167;

Best Local Similarity 75.0%; Pred. No. 65;

Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10 GTCAAGTCTAGTCTTTAGGCCCAAACTAGGTCAAAGG 49

Db 154 GTGAAGTCTAGTCTTTAGGCCCAAACTAGGTCAAAGG 115

# RESULT 14

BE483320/c

LOCUS

DEFINITION 189452 BARC 5BOV Bos taurus cDNA 5', mRNA linear EST 28-AUG-2000

ACCESSION BE483320

VERSION BE483320.1 GI:9602853

KEYWORDS EST.

SOURCE cow.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

REFERENCE  
AUTHORS

1 (bases 1 to 285)  
Wells, K.D.  
Sonstegard, T.S., Capuco, A.V., Van Tassel, C.P., Ashwell, M.S. and  
gland cDNA library

TITLE  
JOURNAL

Mapping of Expressed Sequence Tags from a normalized bovine mammary  
unpublished (2000)

## COMMENT

Contact: Sonstegard TS  
USDA, ARS, Beltsville Agricultural Research Center  
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA  
Tel: 301 504 8416  
Fax: 301 504 8414

Email: tads@anri.barc.usda.gov  
Single pass sequencing Bases called and alt trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.

## PCR Primers

FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCACGACG

Plate: 13 row: O column: 12

Seq primer: ATTTAGGTGACACTATAG.

## FEATURES

source

1..285  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="BARC 5BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
Library made from pooled mRNA isolated from mammary  
tissues at eight physiological, developmental, and disease  
states."

BASE COUNT 72 a 66 c 73 g 74 t

## ORIGIN

Query Match 46.2%; Score 24; DB 10; Length 285;  
Best Local Similarity 75.0%; Pred.No. 78;  
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10 GTCAAAGGTGTCATGCTTTAGGCCCAAACTAGGTCAAAGG 49

Db 152 GTGTAAGATCATGCTTTTAAAGCACAAAATAGGACAAAG 113

## RESULT 15

BF70175/c

## LOCUS

283694 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

## DEFINITION

BF70175

## ACCESSION

BF70175.1

## VERSION

EST.

## KEYWORDS

cow.

## SOURCE

ORGANISM

Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

## REFERENCE

1 (bases 1 to 338)

## AUTHORS

Smith, P.F.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,  
Cavas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett  
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,  
Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and  
Keele, J.W.

## TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle

## JOURNAL

Genome Res. 11 (4), 626-630 (2001)

## MEDLINE

21180013

## COMMENT

Contact: Smith TFL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.

## PCR Primers

FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCACGACG

Plate: 80 row: P column: 1

Seq primer: ATTTAGGTGACACTATAG.

## FEATURES

source

1..338  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 3BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
Library made from pooled tissue from marrow, alveolar  
macrophage, ovary, fetal semitendinosus muscle, and fetal  
longissimus muscle."

BASE COUNT 81 a 75 c 93 g 89 t

## ORIGIN

Query Match 46.2%; Score 24; DB 12; Length 338;  
Best Local Similarity 75.0%; Pred.No. 83;  
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10 GTCAAAGGTGTCATGCTTTAGGCCCAAACTAGGTCAAAGG 49

Db 178 GTGTAAGATCATGCTTTTAAAGCACAAAATAGGACAAAG 139

Search completed: February 20, 2003, 01:08:54

Job time : 256.018 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

DN nucleic - nucleic search, using sw model  
Run on: February 19, 2003, 19:28:33 ; Search time 842.845 seconds  
(without alignments)  
9357.426 Million cell updates/sec  
Title: US-09-808-388-5  
Perfect score: 271  
Sequence: 1 cggcgcaaaactcgtgaaa.....caactcttgagtcctctgag 271

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pi.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_man.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_tgo\_hum.\*  
40: em\_tgo\_mus.\*  
41: em\_trgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	271	100.0	271	6	AX251577	AX251577 Sequence
2	271	100.0	332	6	AX351578	AX351578 Sequence
3	267.4	98.7	1116	6	AX015387	AX015387 Sequence
C 4	267.4	98.7	1268	6	AX015532	AX015532 Sequence
C 5	267.4	98.7	194247	2	AL358253	AL358253 Homo sapi
6	212	78.2	1080	6	IO9231	IO9231 Sequence 36
7	212	78.2	1080	9	HUMRASP1	M22429 Human RASP-
8	170	62.7	157470	2	AL360079	AL360079 Homo sapi
C 9	164.2	60.6	5356	6	AX348874	AX348874 Sequence
C 10	164.2	60.6	5728	6	AX344463	AX344463 Sequence
C 11	164.2	60.6	6083	6	AX251472	AX251472 Sequence
12	150.2	55.4	5356	6	AX348873	AX348873 Sequence
13	150.2	55.4	5728	6	AX344468	AX344468 Sequence
14	150.2	55.4	6083	6	AX251471	AX251471 Sequence
15	118.8	43.8	997	9	BC005919	BC005919 Homo sapi
16	84.4	31.1	3330	6	AX401819	AX401819 Sequence
17	84.4	31.1	3330	10	RNP1A2G	X51523 Rat gene fo
18	83.2	30.7	2697	10	AF375595	AF375595 Rattus no
19	82.8	30.6	3366	10	RATGIIIPHOS	M37127 Rat group I
20	60.6	22.4	497	10	RNP1A2	X52613 Rat DNA for
21	56.6	20.9	157317	2	AC118094	AC118094 Rattus no
22	49.6	18.3	4438	10	MMU32313	U32313 Mus musculu
C 23	49.6	18.3	41125	10	AC002108	AC002108 Gencmic s
24	49.6	18.3	186589	2	AL844178	AL844178 Mus muscu
C 25	39.8	14.7	182209	9	AL357060	AL357060 Human DNA
26	37.8	13.9	125020	9	AF429315	AF429315 Homo sapi
27	37	13.7	183647	2	AC111113	AC111113 Mus muscu
28	36.2	13.4	170466	2	AC103170	AC103170 Rattus no
29	35.6	13.1	206407	2	AC093984	AC093984 Rattus no
30	34.6	12.8	34015	2	AL672255_5	Continuation (6 of
31	34.6	12.8	173424	2	AC110977	AC110977 Rattus no
32	34.4	12.7	150366	2	AC119084	AC119084 Sus scrof
33	34.2	12.6	158073	9	AC093756	AC093756 Homo sapi
34	34.2	12.6	170610	2	AC023935	AC023935 Homo sapi
C 35	34.2	12.6	256657	2	AC129177	AC129177 Mus muscu
36	34	12.5	96660	9	AL442123	AL442123 Human DNA
37	34	12.5	164647	2	AC128714	AC128714 Homo sapi
38	34	12.5	168927	2	AC078797	AC078797 Homo sapi
C 39	34	12.5	171981	3	AC008344	AC008344 Mouse DNA
40	34	12.5	176814	10	AL627444	AL627444 Homo sapi
41	34	12.5	204938	2	AC117437	AC117437 Homo sapi
C 42	34	12.5	277421	3	AE003796	AE003796 Drosophil
43	33.8	12.5	2136	6	AX358609	AX358609 Sequence
44	33.8	12.5	3082	9	AF182218	AF182218 Homo sapi
45	33.8	12.5	3320	6	AX358607	AX358607 Sequence

ALIGNMENTS

RESULT 1  
AX251577 AX251577 271 bp DNA linear PAT 05-OCT-2001  
LOCUS Sequence 5 from Patent WO0168845.  
DEFINITION AX251577  
ACCESSION AX251577  
VERSION AX251577.1 GI:15985000  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequences.  
REFERENCE 1 (bases 1 to 271)  
AUTHORS Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.  
TITLE Inflammation-inducible hybrid promoters, vectors containing same  
and uses thereof  
JOURNAL Patent: WO 0168845-A 5 20-SEP-2001;

Pred. No. is the number of results predicted by chance to have a

## Aventis Pharma S.A. (FR)

FEATURES  
source  
1. .271  
/organism="synthetic construct"  
/db xref="taxon:32630"  
/note="fragment du promoteur PLA2s"  
BASE COUNT 70 a 79 c 71 g 51 t  
ORIGIN

Query Match 100.0%; Score 271; DB 6; Length 271;  
Best Local Similarity 100.0%; Pred. No. 2.5e-75;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGCAAAAGCTGCTGAATGCTTTTGGCATCAGTACTGACAGTAAAGTTTCCCAA 60  
Db 1 CGCGCAAAAGCTGCTGAATGCTTTTGGCATCAGTACTGACAGTAAAGTTTCCCAA 60

Qy 61 TCCTCAACTCTGCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGG 120  
Db 61 TCCTCAACTCTGCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGG 120

Qy 121 CGACAAATCTGAGTCCACCAATGACACAGCCCATCCAGCCTTGCTCCTCACCTACC 180  
Db 121 CGACAAATCTGAGTCCACCAATGACACAGCCCATCCAGCCTTGCTCCTCACCTACC 180

Qy 181 CCCAACTCCAGAGGAGCAGCTATTATTAAGGGAGCAGAGTGCAGAACAAACAGAGC 240  
Db 181 CCCAACTCCAGAGGAGCAGCTATTATTAAGGGAGCAGAGTGCAGAACAAACAGAGC 240

Qy 241 GCCTGGGATACAACTCTGGAGTCTCTGAG 271  
Db 241 GCCTGGGATACAACTCTGGAGTCTCTGAG 271

RESULT 2  
AX251578 AX251578 332 bp DNA linear PAT 05-OCT-2001  
LOCUS  
DEFINITION Sequence 6 from Patent WO0168845.  
ACCESSION AX251578  
VERSION AX251578.1 GI:15985001

KEYWORDS  
synthetic construct.  
synthetic construct.  
artificial sequences.  
REFERENCE 1 (bases 1 to 332)  
AUTHORS Massaad,C., Betenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.  
TITLE Inflammation-inducible hybrid promoters, vectors containing same and uses thereof

JOURNAL Patent: WO 0168845-A 6 20-SEP-2001;

Aventis Pharma S.A. (FR)

FEATURES  
source  
1. .332  
/organism="synthetic construct"  
/db xref="taxon:32630"  
/note="promoteur hybride PP/PLA2s"  
BASE COUNT 96 a 91 c 82 g 63 t  
ORIGIN

Query Match 100.0%; Score 271; DB 6; Length 332;  
Best Local Similarity 100.0%; Pred. No. 2.4e-75;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGCAAAAGCTGCTGAATGCTTTTGGCATCAGTACTGACAGTAAAGTTTCCCAA 60  
Db 62 CGCGCAAAAGCTGCTGAATGCTTTTGGCATCAGTACTGACAGTAAAGTTTCCCAA 121

Qy 61 TCCTCAACTCTGCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGG 120  
Db 122 TCCTCAACTCTGCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGG 181

Qy 121 CGACAAATCTGAGTCCACCAATGACACAGCCCATCCAGCCTTGCTCCTCACCTACC 180  
Db 182 CGACAAATCTGAGTCCACCAATGACACAGCCCATCCAGCCTTGCTCCTCACCTACC 241

Qy 181 CCACACTCCAGAGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAGACG 240  
Db 242 CCACACTCCAGAGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAGACG 301

Qy 241 GCCTGGGATACAACTCTGGAGTCTCTGAG 271  
Db 302 GCCTGGGATACAACTCTGGAGTCTCTGAG 332

RESULT 3  
AX015387 AX015387 1116 bp DNA linear PAT 07-SEP-2000  
LOCUS  
DEFINITION Sequence 53 from Patent WO9951727.  
ACCESSION AX015387  
VERSION AX015387.1 GI:10041367

KEYWORDS  
human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1116)  
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and Pilarsky,C.

TITLE Human nucleic acid sequences of normal ovary tissue  
JOURNAL Patent: WO 9951727-A 53 14-OCT-1999;  
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN  
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN  
(DE); PILARSKY CHRISTIAN (DE)

FEATURES  
Location/Qualifiers  
source  
1. .1116  
/organism="Homo sapiens"  
/db xref="taxon:9606"  
BASE COUNT 311 a 311 c 266 g 228 t  
ORIGIN

Query Match 98.7%; Score 267.4; DB 6; Length 1116;  
Best Local Similarity 99.6%; Pred. No. 3.3e-74;  
Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CGGCAAACTGCCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGTTTCCCAATC 62  
Db 6 CTGCAAACTGCCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGTTTCCCAATC 65

Qy 63 CTCAACTCTCTGCTGCAGCTGATGAGGGAGGAGGATTAAGGGGTATGGGCG 122  
Db 66 CTCAACTCTCTGCTGCAGCTGATGAGGGAGGAGGATTAAGGGGTATGGGCG 125

Qy 123 ACCAATCTGAGTCCCAACTGACCAAGCCCATCCAGCCTTGTGCTCCTACCTACCCC 182  
Db 126 ACCAATCTGAGTCCCAACTGACCAAGCCCATCCAGCCTTGTGCTCCTACCTACCCC 185

Qy 183 CAACCTCCAGAGGAGCAGCTATTATTAAGGGAGGAGGATGAGGAGGAGGAGGAGGAG 242  
Db 186 CAACCTCCAGAGGAGCAGCTATTATTAAGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 245

Qy 243 CTGGGATACAACTCTGGAGTCTCTGAG 271  
Db 246 CTGGGATACAACTCTGGAGTCTCTGAG 274

RESULT 4  
AX015532/c AX015532 1268 bp DNA linear PAT 07-SEP-2000  
LOCUS  
DEFINITION Sequence 226 from Patent WO9951727.  
ACCESSION AX015532  
VERSION AX015532.1 GI:10041414

KEYWORDS  
human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1268)

AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and Pilarsky,C.  
 TITLE Human nucleic acid sequences of normal ovary tissue  
 JOURNAL Patent: WO 9951727-A 226 14-OCT-1999;  
 SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN  
 BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN  
 (DE); PILARSKY CHRISTIAN (DE)  
 FEATURES Location/Qualifiers  
 source 1..1268  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 ASE COUNT 262 a 313 c 343 g 350 t  
 RIGIN  
 Query Match 98.7%; Score 267.4; DB 6; Length 1268;  
 Best Local Similarity 99.6%; Pred. No. 3.3e-74;  
 Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Y 3 CGGCAAACTGCTGAAATGTGTTTGGCATCAGTACTGACACGTAAGGTTTCCCAATC 62  
 b 1263 CTGCAAACTGCCTGAAATGTGTTTGGCATCAGTACTGACACGTAAGGTTTCCCAATC 1204  
 Y 63 CTCAACTCTGTCCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGGCG 122  
 b 1203 CTCAACTCTGTCCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGGCG 1144  
 Y 123 ACCAATCTGAGTCCCACTGACACGCGCCATCCCGACCTTGTGCTCCTACCTACCC 182  
 b 1143 ACCAATCTGAGTCCCACTGACACGCGCCATCCCGACCTTGTGCTCCTACCTACCC 1084  
 Y 183 CAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCGAAGCAAAACAAGACGC 242  
 b 1083 CAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCGAAGCAAAACAAGACGC 1024  
 Y 243 CTGGGGATACAACTCTGGAGTCTCTGAG 271  
 b 1023 CTGGGGATACAACTCTGGAGTCTCTGAG 995  
 RESULT 5  
 AL358253/c 194247 bp DNA linear HTG 25-JUL-2002  
 JOCUS  
 DEFINITION Homo sapiens chromosome 1 clone RP11-460G22, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, in ordered pieces.  
 ACCESSION AL358253  
 VERSION AL358253.14 GI:21998163  
 HTG; HTGS PHASE2; HTGS ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS Homo sapiens.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 194247)  
 Bagguley,C.  
 Direct Submission  
 TITLE Submitted (24-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Jul 26, 2002 this sequence version replaced gi:21748207.  
 COMMENT  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 ----- Project Information  
 Center project name: BA460G22  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator; 0% of reads  
 Chemistry: Dye-primer-amersham; 1% of reads  
 Chemistry: Dye-terminator ET-amersham; 3% of reads  
 Dye-terminator Big Dye; 95% of reads  
 Consensus quality: 194218 bases at least Q40  
 Consensus quality: 194247 bases at least Q30

Consensus quality: 194247 bases at least Q20  
 Insert size: 194247; sum-of-contigs  
 Insert size: 187546; 12.3% error; agarose-fp  
 Quality coverage: 10.76x in Q20 bases; sum-of-contigs Quality  
 coverage: 12.13x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 FEATURES Location/Qualifiers  
 source 1..194247  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-460G22"  
 /clone\_lib="RPC1-11.2"  
 misc\_feature 1..194247  
 /note="assembly fragment:05625"  
 BASE COUNT 52994 a 42433 c 43263 g 55557 t  
 ORIGIN  
 Query Match 98.7%; Score 267.4; DB 2; Length 194247;  
 Best Local Similarity 99.6%; Pred. No. 2.9e-74;  
 Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 CGGCAAACTGCTGAAATGTGTTTGGCATCAGTACTGACACGTAAGGTTTCCCAATC 62  
 Db 165725 CTGCAAACTGCTGAAATGTGTTTGGCATCAGTACTGACACGTAAGGTTTCCCAATC 165666  
 QY 63 CTCAACTCTGTCCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGGCG 122  
 Db 165665 CTCAACTCTGTCCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGGCG 165606  
 QY 123 ACCAATCTGAGTCCCACTGACACGCGCCATCCCGACCTTGTGCTCCTACCTACCC 182  
 Db 165605 ACCAATCTGAGTCCCACTGACACGCGCCATCCCGACCTTGTGCTCCTACCTACCC 165546  
 QY 183 CAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCGAAGCAAAACAAGACGC 242  
 Db 165545 CAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCGAAGCAAAACAAGACGC 165486  
 QY 243 CTGGGGATACAACTCTGGAGTCTCTGAG 271  
 Db 165485 CTGGGGATACAACTCTGGAGTCTCTGAG 165457  
 RESULT 6  
 I09231 109231 1080 bp DNA linear PAT 02-DEC-1994  
 DEFINITION Sequence 36 from Patent WO 8901773.  
 ACCESSION I09231  
 VERSION I09231.1 GI:588062  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 Unclassified.  
 1 (bases 1 to 1080)  
 REFERENCE Johnson,L.K., Seilhamer,J.J., Pruzanski,W. and Vadas,P.  
 AUTHORS SYNOPSIS PHOSPHOLIPASES  
 JOURNAL Patent: WO 8901773-A 36 09-MAR-1989;  
 FEATURES Location/Qualifiers  
 source 1..1080  
 /organism="unknown"  
 BASE COUNT 306 a 242 c 303 g 229 t  
 ORIGIN  
 Query Match 78.2%; Score 212; DB 6; Length 1080;  
 Best Local Similarity 93.1%; Pred. No. 1.6e-56;  
 Matches 255; Conservative 0; Mismatches 15; Indels 4; Gaps 3;  
 QY 1 CGCGGCAAACTGCTGAAATGTGTTTGGCATCAGTACTGACACGTAAGGTTTCCCAATC 59

Db 763 CTGGCAAACTGCCTGAATGTGTTTGGCATCAGCTACTGACACGTAAGGGTTTCCCA 822

Qy 60 ATCTCAACTCTCTCTG--CCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTAT 117

Db 823 ATCTCAACTCTCTCTGCTGCCAGCTGATGAGGGGAGAAAGGATTACCTAGGGGTAT 882

Qy 118 GGGGACCAATCTGATGTCACCACTGACCGCCATCCCGACCTTGCTGCTCACCT 177

Db 883 GGGGACCAATCTGATGTCACCACTGACCGCCATCCCGACCTTGCTGCTCACCT 942

Qy 178 ACCCCCAACTCCAGAGGGAGCAGCTATTATAGGGGAGCAGAGTGCGAACAAACAG 237

Db 943 ACCCCCAACT--CAGAGGGAGCAGCTATTATAGGGGAGCAGAGTGCGAACAAACAG 1001

Qy 238 ACGCCTGGGGATACAACTCTGGAGTCTCTGAG 271

Db 1002 ACGCCTGGGGATACAACTCTGGAGTCTCTGAG 1035

RESULT 7

HUMRASFA1

LOCUS Human RASFA1 1080 bp DNA linear PRI 27-APR-1993

DEFINITION Human RASFA1 A PIA2 gene encoding synovial phospholipase A-2, exon 1.

ACCESSION M22429 J04704

VERSION M22429.1 GI:190884

KEYWORDS synovial phospholipase A-2; synovial phospholipase A-2-peak A.

SEGMENT 1 of 2

SOURCE Human DNA.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1080)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

JOURNAL Seilhamer, J.J.

REFERENCE 2 (bases 561 to 1080)

AUTHORS Seilhamer, J.J., Pruzanski, W., Vadas, P., Plant, S., Miller, J.A., Kloss, J., and Johnson, L.K.

TITLE Cloning and recombinant expression of phospholipase A2 present in rheumatoid arthritic synovial fluid

JOURNAL J. Biol. Chem. 264 (10), 5335-5338 (1989)

MEDLINE 89174566

PUBMED 2925608

COMMENT Draft entry and computer-readable sequence [2] kindly submitted by J.J. Seilhamer 07-FEB-1989.

FEATURES

source

1. 1080

/organism="Homo sapiens"

/db\_xref="taxon:9606"

intron

1035..>1080

/note="synovial phospholipase, intron A"

BASE COUNT 306 a 242 c 303 g 229 t

ORIGIN Unreported.

Query Match 78.2%; Score 212; DB 9; Length 1080;

Best Local Similarity 93.1%; Pred. No. 1.6e-56;

Matches 255; Conservative 0; Mismatches 15; Indels 4; Gaps 3;

Qy 1 CGCGCAAACTGCCTGAATGTGTTTGGCATCAGCTACTGACACGTAAGGGTTTCCCA 59

Db 763 CTGGCAAACTGCCTGAATGTGTTTGGCATCAGCTACTGACACGTAAGGGTTTCCCA 822

Qy 60 ATCTCAACTCTCTCTG--CCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTAT 117

Db 823 ATCTCAACTCTCTCTGCTGCCAGCTGATGAGGGGAGAAAGGATTACCTAGGGGTAT 882

Qy 118 GGGGACCAATCTGATGTCACCACTGACCGCCATCCCGACCTTGCTGCTCACCT 177

Db 883 GGGGACCAATCTGATGTCACCACTGACCGCCATCCCGACCTTGCTGCTCACCT 942

Qy 178 ACCCCCAACTCCAGAGGGAGCAGCTATTATAGGGGAGCAGAGTGCGAACAAACAG 237

Db 943 ACCCCCAACT--CAGAGGGAGCAGCTATTATAGGGGAGCAGAGTGCGAACAAACAG 1001

Qy 238 ACGCCTGGGGATACAACTCTGGAGTCTCTGAG 271

Db 1002 ACGCCTGGGGATACAACTCTGGAGTCTCTGAG 1035

RESULT 8

AL360079/c

LOCUS Homo sapiens chromosome 1 clone RP11-66M4, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 22 unordered pieces.

DEFINITION

ACCESSION AL360079

VERSION AL360079.3 GI:9801103

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.

SOURCE human

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 157470)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

JOURNAL Mclay, K.

REFERENCE Direct Submission

TITLE Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT requests: clonerequest@sanger.ac.uk

On Aug 14, 2000 this sequence version replaced gi:8919533.

----- Sanger Centre

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BAC6M4

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator ET-amersham; 1% of reads Chemistry:

Dye-terminator Big Dye; 98% of reads

Consensus quality: 145607 bases at least Q40

Consensus quality: 150854 bases at least Q30

Consensus quality: 153362 bases at least Q20

Insert size: 155370; sum-of-coverage

Quality coverage: 175968; 2.3% error; agarose-fp

Quality coverage: 3.40x in Q20 bases; sum-of-coverage Quality coverage: 3.16x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 15108: contig of 15108 bp in length

15109 15208: gap of 100 bp

15209 20077: contig of 4869 bp in length

20078 20177: gap of 100 bp

20178 26598: contig of 6421 bp in length

26599 26698: gap of 100 bp

26699 31874: contig of 5176 bp in length

31875 31974: gap of 100 bp

31975 39580: contig of 7606 bp in length

39581 39680: gap of 100 bp

39681 62977: contig of 23297 bp in length

62978 63077: gap of 100 bp

63078 75654: contig of 12577 bp in length

75655 75754: gap of 100 bp

75755 90667: contig of 14913 bp in length

90668 90767: gap of 100 bp

90768 93945: contig of 3178 bp in length

93946 94045: gap of 100 bp

94046 99022: contig of 4977 bp in length

99023 99122: gap of 100 bp

99123 101699: contig of 2577 bp in length

<hr/>					
misc_feature      144605..157470					
BASE COUNT	43730	a	34140	c	34257 g    43227 t    2116 others
ORIGIN					
<hr/>					
Query Match	62.7%	Score	170;	DB 2;	Length 157470;
Best Local Similarity	97.2%;	Pred.	No. 3.5e-43;		
Matches	173;	Conservative	0;	Mismatches	5; Indels 0; Gaps 0;
<hr/>					
QY	3	CGCAAAACTGCTGAATGTGTTTGGCATCAGCTACTGCACAGTAAGTTTCCCAATC	62		
Db	15386	CTGCAAACTGCTGAATGTGTTTGGCATCAGCTACTGCACAGTAAGTTTCCCAATC	15327		
<hr/>					
QY	63	CTCAACTGTCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGCG	122		
Db	15326	CTCACCTGTCTGCCAGCTAATGAGGGGAAGAAAGGATTACCTAGGGGTATGGCG	15267		
<hr/>					
QY	123	ACCATTCCTGAGTGACCAACTGACACAGCGCCATCCCAGCCTTGTGCTCACCCTACC	180		
Db	15266	ACCATTCTTGAGTGACCAACTGACACAGCGCCATCCCAGCCTTGTGCTCACCCTACC	15209		
<hr/>					
RESULT 9					
AX348874/c	AX348874	Sequence 332 from Patent WO0202807.	5356 bp	DNA	linear PAT 06-FEB-2002
LOCUS	AX348874	Accession	AX348874		
DEFINITION	AX348874.1	GI:18614909			
VERSION	AX348874.1	GI:18614909			
KEYWORDS	synthetic construct.				
SOURCE	synthetic construct				
ORGANISM	artificial sequences.				
REFERENCE	1				
AUTHORS	Olek,A., Piepenbrock,C. and Berlin,K.				
TITLE	Diagnosis of diseases associated with cell signalling				
JOURNAL	Patent: WO 0202807-A 332 10-JAN-2002;				
EPIGENOMICS AG (DE)					
FEATURES	Location/Qualifiers				
source	1..5356				
/organism="synthetic construct"					
/db xref="taxon:32630"					
/note="Chemically treated genomic DNA (Homo sap.)ens)"					
BASE COUNT	1357	a	60 c	1167 g	2772 t
ORIGIN					
<hr/>					
Query Match	60.6%	Score	164.2;	DB 6;	Length 5356;
Best Local Similarity	76.2%;	Pred.	No. 2.7e-41;		
Matches	202;	Conservative	0;	Mismatches	63; Indels 0; Gaps 0;
<hr/>					
QY	6	CAAAAGCTGCTGAATGTGTTTGGCATCAGCTACTGCACAGTAAGTTTCCCAATCCTC	65		
Db	1324	CAAACTACTATAATATATTTTAACATCACTACTAACAGTAATTTCCCAATCCTC	1265		
<hr/>					
QY	66	AACCTGTCTCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGGGAC	125		
Db	1264	AACCTGTCTCTGCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGGGAC	1205		
<hr/>					
QY	126	AATCTTGAGTGACCAACTGACCGCCATCCCCAGCCTTGTGCTCACCCTACCCCAA	185		
Db	1204	AATCTTAAATGACCAACTAACCGCCATCCCCAACCCTTATACCTCACCCTACCCCAA	1145		
<hr/>					
QY	186	CTTCCCAGAGGAGCAGCTATTATTAGGGGAGCAGAGTGCAGACAACAAGACGGCCTG	245		
Db	1144	CTTCCCAGAGGAGCAGCTATTATTAGGGGAGCAGAGTGCAGACAACAAGACGGCCTG	1085		
<hr/>					
QY	246	GGGATCAACTCTGGAGTCTCTGA	270		
Db	1084	AAAAATCAACTCTAAAATCCTCTAA	1060		
<hr/>					
RESULT 10					
AX344469/c	AX344469	Sequence 332 from Patent WO0202807.	5728 bp	DNA	linear PAT 01-FEB-2002
LOCUS	AX344469	Accession	AX344469		
DEFINITION	AX344469.1	GI:18614909			
VERSION	AX344469.1	GI:18614909			
KEYWORDS	synthetic construct.				
SOURCE	synthetic construct				
ORGANISM	artificial sequences.				
REFERENCE	1				
AUTHORS	Olek,A., Piepenbrock,C. and Berlin,K.				
TITLE	Diagnosis of diseases associated with cell signalling				
JOURNAL	Patent: WO 0202807-A 332 10-JAN-2002;				
EPIGENOMICS AG (DE)					
FEATURES	Location/Qualifiers				
source	1..5356				
/organism="synthetic construct"					
/db xref="taxon:32630"					
/note="Chemically treated genomic DNA (Homo sap.)ens)"					
BASE COUNT	1357	a	60 c	1167 g	2772 t
ORIGIN					
<hr/>					

```

DEFINITION      Sequence 316 from Patent WO0200926.
ACCESSION       AX344469
VERSION         AX344469.1 GI:18492357
KEYWORDS        synthetic construct.
SOURCE          synthetic construct.
ORGANISM        artificial sequences.
REFERENCE       1
AUTHORS         Olek, A., Piepenbrock, C. and Berlin, K.
TITLE          Diagnosis of diseases associated with signal transduction
JOURNAL         Patent: WO 0200926-A 316 03-JAN-2002;
Epigenomics AG (DE)
FEATURES        Location/Qualifiers
                1..5728
                /organism="synthetic construct"
                /db xref="taxon:32630"
                /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     1462 a 79 c 1271 g 2916 t
ORIGIN

Query Match      60.6%; Score 164.2; DB 6; Length 5728;
Best Local Similarity 76.2%; Pred. No. 2.7e-41;
Matches 202; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 6 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 65
Db 6 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 65

Qy 969 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 910
Db 969 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 910

Qy 66 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 125
Db 66 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 125

Qy 909 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 850
Db 909 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 850

Qy 126 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 185
Db 126 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 185

Qy 849 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 790
Db 849 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 790

Qy 186 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 245
Db 186 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 245

Qy 789 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 730
Db 789 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 730

Qy 246 GGGATACAACTCTGGAGTCTCTGA 270
Db 246 GGGATACAACTCTGGAGTCTCTGA 270

Qy 729 AAATACAACTCTAAATCTCTTAA 705
Db 729 AAATACAACTCTAAATCTCTTAA 705

RESULT 11
LOCUS           AX251472/c
DEFINITION      Sequence 440 from Patent WO0168912.
ACCESSION       AX251472
VERSION         AX251472.1 GI:15984895
KEYWORDS        synthetic construct.
SOURCE          synthetic construct.
ORGANISM        artificial sequences.
REFERENCE       1 (bases 1 to 6083)
AUTHORS         Olek, A., Piepenbrock, C. and Berlin, K.
TITLE          Diagnosis of diseases associated with tumor suppressor genes and
                oncogenes
JOURNAL         Patent: WO 0168912-A 440 20-SEP-2001;
Epigenomics AG (DE)
FEATURES        Location/Qualifiers
                1..6083
                /organism="synthetic construct"
                /db xref="taxon:32630"
                /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     1525 a 83 c 1367 g 3108 t
ORIGIN

Query Match      60.6%; Score 164.2; DB 6; Length 6083;
Best Local Similarity 76.2%; Pred. No. 2.7e-41;
Matches 202; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 6 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 65
Db 6 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 65

Qy 969 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 910
Db 969 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 910

Qy 66 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 125
Db 66 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 125

Qy 909 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 850
Db 909 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 850

Qy 126 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 185
Db 126 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 185

Qy 849 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 790
Db 849 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 790

Qy 186 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 245
Db 186 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 245

Qy 789 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 730
Db 789 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 730

Qy 246 GGGATACAACTCTGGAGTCTCTGA 270
Db 246 GGGATACAACTCTGGAGTCTCTGA 270

Qy 729 AAATACAACTCTAAATCTCTTAA 705
Db 729 AAATACAACTCTAAATCTCTTAA 705

RESULT 12
LOCUS           AX348873
DEFINITION      Sequence 331 from Patent WO0202807.
ACCESSION       AX348873
VERSION         AX348873.1 GI:18614908
KEYWORDS        synthetic construct.
SOURCE          synthetic construct.
ORGANISM        artificial sequences.
REFERENCE       1
AUTHORS         Olek, A., Piepenbrock, C. and Berlin, K.
TITLE          Diagnosis of diseases associated with cell signalling
JOURNAL         Patent: WO 0202807-A 331 10-JAN-2002;
Epigenomics AG (DE)
FEATURES        Location/Qualifiers
                1..5356
                /organism="synthetic construct"
                /db xref="taxon:32630"
                /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     1476 a 60 c 1356 g 2464 t
ORIGIN

Query Match      55.4%; Score 150.2; DB 6; Length 5356;
Best Local Similarity 72.7%; Pred. No. 7.9e-37;
Matches 194; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 5 GCAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 64
Db 5 GCAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 64

Qy 4032 GTAAATTTGTTGAAATGTTTGGTATTGATTGATTGATTGATTGATTGATTGATTGATT 4091
Db 4032 GTAAATTTGTTGAAATGTTTGGTATTGATTGATTGATTGATTGATTGATTGATTGATT 4091

Qy 65 CAACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 124
Db 65 CAACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 124

Qy 4092 TAATTTTGTGTTTGTAGTTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 4151
Db 4092 TAATTTTGTGTTTGTAGTTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 4151

Qy 125 CAATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 184
Db 125 CAATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 184

Qy 4152 TAATTTGAGTATTAAATGATTACGTTTATTTATTTGTTTGTGTTTATTTATTTT 4211
Db 4152 TAATTTGAGTATTAAATGATTACGTTTATTTATTTGTTTGTGTTTATTTATTTT 4211

Qy 185 ACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTC 244
Db 185 ACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTC 244

Qy 4212 ATTTTGTAGAGGAGTAGTTTATTTAAGGGGAGTAGGAGTGTAGATTAATAAATAGACGGTTT 4271
Db 4212 ATTTTGTAGAGGAGTAGTTTATTTAAGGGGAGTAGGAGTGTAGATTAATAAATAGACGGTTT 4271

Qy 245 GGGATACAACTCTGGAGTCTCTGA 271
Db 245 GGGATACAACTCTGGAGTCTCTGA 271

Qy 4272 GGGGATATAATTTTGGAGTCTTTTGGAG 4298
Db 4272 GGGGATATAATTTTGGAGTCTTTTGGAG 4298

RESULT 13
LOCUS           AX344468
DEFINITION      Sequence 315 from Patent WO0200926.

```

```

Qy 6 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 65
Db 1324 CAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 1265

Qy 66 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 125
Db 1264 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 1205

Qy 126 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 185
Db 1204 AATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 1145

Qy 186 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 245
Db 1144 CTCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTG 1085

Qy 246 GGGATACAACTCTGGAGTCTCTGA 270
Db 1084 AAAATACAACTCTAAAATCTCTTAA 1060

RESULT 12
LOCUS           AX348873
DEFINITION      Sequence 331 from Patent WO0202807.
ACCESSION       AX348873
VERSION         AX348873.1 GI:18614908
KEYWORDS        synthetic construct.
SOURCE          synthetic construct.
ORGANISM        artificial sequences.
REFERENCE       1
AUTHORS         Olek, A., Piepenbrock, C. and Berlin, K.
TITLE          Diagnosis of diseases associated with cell signalling
JOURNAL         Patent: WO 0202807-A 331 10-JAN-2002;
Epigenomics AG (DE)
FEATURES        Location/Qualifiers
                1..5356
                /organism="synthetic construct"
                /db xref="taxon:32630"
                /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     1476 a 60 c 1356 g 2464 t
ORIGIN

Query Match      55.4%; Score 150.2; DB 6; Length 5356;
Best Local Similarity 72.7%; Pred. No. 7.9e-37;
Matches 194; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 5 GCAAACTGCCTGAAATGTTTGGCATCAGCTACTGACAGTAAGGTTTCCCAATCCTC 64
Db 4032 GTAAATTTGTTGAAATGTTTGGTATTGATTGATTGATTGATTGATTGATTGATTGATT 4091

Qy 65 CAACTCTGTCTGCCAGCTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 124
Db 4092 TAATTTTGTGTTTGTAGTTGATGAGGGGAAGGAGGATTAACCTAGGGGTATGGCGGACC 4151

Qy 125 CAATCTGTAGTCCACCACTGACAGCGCCATCCCGCTTGTGCTTACCTACCTACCCCAA 184
Db 4152 TAATTTGAGTATTAAATGATTACGTTTATTTATTTGTTTGTGTTTATTTATTTT 4211

Qy 185 ACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGCAACAAACAGCGCCTC 244
Db 4212 ATTTTGTAGAGGAGTAGTTTATTTAAGGGGAGTAGGAGTGTAGATTAATAAATAGACGGTTT 4271

Qy 245 GGGATACAACTCTGGAGTCTCTGA 271
Db 4272 GGGGATATAATTTTGGAGTCTTTTGGAG 4298

RESULT 13
LOCUS           AX344468
DEFINITION      Sequence 315 from Patent WO0200926.

```

[illegible][illegible]

```
/translation="MKTLLLLAVIMIFGLLQAHGNLVNPHRMIKLTTGKEAALSVGFY
GCHCGVGRGSPKDATDRCCVTHDCCYKRLEKRGCGTKFLSYKFPNSGSRITCAKQDS
CRSOLCECDKAAATCFARNKTTYKNKYOYVSNKHCRCGSTPRC"
```

BASE COUNT	289 a	279 c	231 g	198 t
ORIGIN				

Query Match 43.8%; Score 118.8; DB 9; Length 997;  
Best Local Similarity 83.3%; Pred. No. 8,6e-27;  
Matches 135; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 110 AGGGGTATGGGCGACCAATCTGTAGTCCACCAACTGACCAAGGCCATCCCCAGCCTTGTG 169

D<sub>b</sub> 3 AGGAAAAGAGCAACAGATCCAGGAGCATTCACCTGCCCTGTCTCCAACAGCCTTGTG 62

QV 170 CCTCACCTACCCCCAACCTCCAGAGGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGAA 229

63 CCTCAGCTACCCGCCAACCCTCCAGAGGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGAA 122

230 CAAACAAGACGGCCTGGGGATACAACCTCTGGAGTCCTCTGAG 271

123 CAAACAAGACGGCGTGGGATACAACTCTGGAGTCTCTGAG 164

Search completed: February 19, 2003, 23:08:17  
Job time : 943.845 secs



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:26:38 ; Search time 164.707 Seconds  
(without alignments)  
3705.323 Million cell updates/sec

File: US-09-808-388-5  
Effect score: 271  
Sequence: 1 cgcggcaaacgcctgaaa.....caactctggagctcctcgag 271

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	271	100.0	271	22	Partial human PLA2
2	271	100.0	332	22	Partial synthetic
3	267.4	98.7	1116	20	Human normal ovary
4	267.4	98.7	6172	10	HindIII fragment o
5	212	78.2	1080	10	Nucleotide sequenc
6	164.2	60.6	5356	24	Chemically treated
7	164.2	60.6	5728	24	Signal transductio
8	164.2	60.6	6083	22	Tumour suppressor
9	150.2	55.4	5356	24	Chemically treated

10	150.2	55.4	5728	24	ABK31472	Signal transductio
11	150.2	55.4	5083	22	AA546714	Tumour suppressor
12	118.8	43.8	204	23	ABV09174	Human prostate exp
13	118.8	43.1	481	23	ABV39325	Human prostate exp
14	84.4	31.1	3330	24	ABK63588	Rat sequence diffe
15	80.8	29.8	1076	21	AA715635	Human prostate can
16	79.2	29.2	1160	23	ABV22259	Human prostate exp
17	79.2	29.2	1160	23	ABV23366	Human prostate exp
18	79.2	29.2	1160	23	ABV23388	Human prostate exp
19	79.2	29.2	1160	23	ABV25535	Human prostate exp
20	79.2	29.2	1160	23	ABV28096	Human prostate exp
21	79.2	29.2	1160	23	ABV29221	Human prostate exp
22	79.2	29.2	1160	23	ABV29244	Human prostate exp
23	78.4	28.9	126	23	ABV14604	Human prostate exp
24	78.4	28.9	165	23	ABV05435	Human prostate exp
25	78.4	28.9	516	23	ABV35682	Human prostate exp
26	78.4	28.9	516	23	ABV4487	Human prostate exp
27	48.2	17.8	119	23	ABV08557	Human prostate exp
28	48.2	17.8	390	23	ABV38459	Human prostate exp
29	37.6	13.9	441	23	ABV38462	Human prostate exp
30	37	13.7	83	23	ABV14206	Human prostate exp
31	37	13.7	119	23	ABV08718	Human prostate exp
32	37	13.7	120	23	ABV08560	Human prostate exp
33	37	13.7	276	23	ABV08936	Human prostate exp
34	37	13.7	387	23	ABV44134	Human prostate exp
35	37	13.7	428	23	ABV35300	Human prostate exp
36	37	13.7	448	23	ABV38616	Human prostate exp
37	37	13.7	450	23	ABV38824	Human prostate exp
38	36	13.3	123	23	ABV05037	Human prostate exp
39	34	12.5	1687	23	ABL12468	Drosophila melanog
40	34	12.5	1687	23	ABL18110	Drosophila melanog
41	34	12.5	1687	23	ABL18210	Drosophila melanog
42	33.8	12.5	420	21	AA661755	cDNA encoding a hu
43	33.8	12.5	1441	21	AA661753	cDNA encoding a hu
44	33.8	12.5	2236	21	AA661749	cDNA encoding a hu
45	33.8	12.5	2604	21	AA661758	cDNA encoding a hu

ALIGNMENTS

RESULT 1  
AAI64307  
AAI64307 standard; DNA; 271 BP.  
XX AAI64307;  
AC AAI64307;  
XX 15-NOV-2001 (first entry)  
DT Partial human PLA2SIIA gene promoter.  
DE  
XX PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2SIIA;  
KW peroxisome proliferator activated receptor; human;  
KW secreted non-pancreatic phospholipase A2; ds.  
XX Homo sapiens.  
OS  
XX  
XX WO200168845-A2.  
XX 20-SEP-2001.  
XX  
XX 14-MAR-2001; 2001WO-FR00759.  
XX  
XX 14-MAR-2000; 2000FR-0003262.  
XX 13-APR-2000; 2000US-0196959.  
XX (AVET ) AVENTIS PHARMA SA.  
XX  
XX Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;  
XX WPI; 2001-582451/65.  
XX



DR WPI; 1999-552352/47.  
 KX Nucleic acid sequences potentially useful in diagnosis or therapy of  
 PT ovarian cancer -  
 KX  
 KX  
 PS Claim 3; Page 164; 274pp; German.  
 KX  
 CC This invention describes novel nucleic acid sequences that are highly  
 CC expressed in normal ovary tissue. Artificial chromosomes and cosmid  
 CC clones containing the sequences can be used as gene transfer vehicles.  
 CC The sequences can be used to produce DNA fragments containing  
 CC full-length genes. Host cells transformed with the sequences can be used  
 CC to produce polypeptides or polypeptide fragments, which can be used to  
 CC screen phase displays for polypeptides that bind to them, or as tools for  
 CC identifying agents active against ovarian cancer, or to prepare  
 CC medicaments for treating ovarian cancer. The cDNA sequences can be used  
 CC to obtain genomic genes, their promoters, enhancers, silencers, exon  
 CC structures, intron structures and their splice variants. AAZ4122-241324  
 CC represent cDNA sequences derived from normal human ovarian tissue and  
 CC which encode the protein fragments represented in AAY59724-Y59837.  
 KX  
 KX Sequence 1116 BP; 311 A; 311 C; 266 G; 228 T; 0 other;

Query Match 98.7%; Score 267.4; DB 20; Length 1116;  
 Best Local Similarity 99.6%; Pred. No. 2.7e-75;  
 Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2Y 3 CGGCAAACTGCCTGAATGTGTTTGGCATCAGCTACTGACACGTAAAGTTTCCCAATC 62  
 Db 6 CTGCAAACTGCTGAATGTGTTTGGCATCAGCTACTGACACGTAAAGTTTCCCAATC 65  
 2Y 63 CTCACCTCTGCTGCCAGCTGATGAGGGGAAGAAAGGATTAACCTAGGGTATGGCG 122  
 Db 65 CTCACCTCTGCTGCCAGCTGATGAGGGGAAGAAAGGATTAACCTAGGGTATGGCG 125  
 2Y 123 ACCATCTGAGTCCACCACTGACACGCCATCCCGAGCTGTGCTTCACTACCCC 182  
 Db 126 ACCATCTGAGTCCACCACTGACACGCCATCCCGAGCTGTGCTTCACTACCCC 185  
 2Y 183 CAACCTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGAGTGCGAACAACAAGACGGC 242  
 Db 186 CAACCTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGAGTGCGAACAACAAGACGGC 245  
 2Y 243 CTGGGGATACACTCTGAGTCCCTCTGAG 271  
 Db 246 CTGGGGATACACTCTGAGTCCCTCTGAG 274

RESULT 4  
 XAN91825  
 ID AAN91825 standard; DNA; 6172 BP.  
 KX  
 AC AAN91825;  
 KX  
 DT 31-JUL-1992 (second entry)  
 KX  
 DE HindIII fragment of PLA2 8.5 EMBL3 encoding human inflammatory  
 DE phospholipase A2.  
 KX  
 KX Inflammation; acid stable; phosphatide 2-acylhydrolase; lipolytic;  
 KX glycerophospholipids; non-pancreatic; ss.  
 KX  
 DS Homo sapiens.

Key Location/Qualifiers  
 FT exon 1..2492  
 FT /\*tag= a  
 FT /number= 1  
 FT 2702..2846  
 FT /\*tag= b  
 FT /number= 2  
 FT 3105..3211  
 FT /\*tag= c

FT exon /number= 3  
 FT 5383..6172  
 FT /\*tag= d  
 FT /number= 4  
 FT 2453..2721  
 FT /\*tag= e  
 FT 5771..5776  
 FT /\*tag= f  
 FT misc\_feature 2715..2826  
 FT /tag= g  
 FT /notes="Claim 24"  
 FT misc\_feature 2715..2826  
 FT /\*tag= g  
 FT /notes="Claim 24"  
 XX W08909818-A.  
 XX  
 XX 19-OCT-1989.  
 PD  
 XX  
 XX 11-APR-1989; 89WO-US01418.  
 XX  
 XX 15-APR-1988; 88US-0181893.  
 XX  
 XX (BIOJ ) BIOGEN INC.  
 XX  
 XX Kramer RM, Peginsky RB, Hession C;  
 FI  
 XX  
 XX WPI: 1989-324225/44.  
 DR P-PSDB; AAP93112.  
 XX  
 XX Acid stable phospholipase A2 - used for prodn. of antibodies and in  
 PT the treatment or diagnosis of inflammation of diseases.  
 XX  
 XX Claim 27; Fig 12; 84pp; English.  
 XX  
 XX A genomic DNA library was prepd. from a mutant fibroblast cell line  
 CC which contains 5 copies of the X chromosome (GM5009). The plaques  
 CC were screened for a gene encoding PLA2 using probes designed from  
 CC peptides derived from the purified protein. A positive clone, PLA2  
 CC 8.5 EMBL3 was purified and a 6.2 kb insert sequenced (shown here).  
 CC Corresponding cDNA sequences (i.e. without introns; see N97209) can  
 CC be ligated into expression vectors for the prodn. of recombinant  
 CC PLA2. The protein, and antibodies raised to it, can be used for  
 CC diagnosis of inflammation and tissue injury associated with various  
 CC diseases.  
 CC See also N91826-33 and N97209.  
 XX  
 XX Sequence 6172 BP; 1624 A; 1408 C; 1740 G; 1400 T; 0 other;  
 SQ  
 Query Match 98.7%; Score 267.4; DB 10; Length 6172;  
 Best Local Similarity 99.6%; Pred. No. 4.9e-75;  
 Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 3 CGGCAAACTGCCTGAATGTGTTTGGCATCAGCTACTGACACGTAAAGTTTCCCAATC 62  
 Db 1378 CTGCAAACTGCCTGAATGTGTTTGGCATCAGCTACTGACACGTAAAGTTTCCCAATC 1437  
 Qy 63 CTCACCTCTGCTGCCAGCTGATGAGGGGAAGAAAGGATTAACCTAGGGTATGGCG 122  
 Db 1438 CTCACCTCTGCTGCCAGCTGATGAGGGGAAGAAAGGATTAACCTAGGGTATGGCG 1497  
 Qy 123 ACCATCTGAGTCCACCACTGACACGCCATCCCGAGCTGTGCTTCACTACCCC 182  
 Db 1498 ACCATCTGAGTCCACCACTGACACGCCATCCCGAGCTGTGCTTCACTACCCC 1557  
 Qy 183 CAACCTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGAGTGCGAACAACAAGACGGC 242  
 Db 1558 CAACCTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGAGTGCGAACAACAAGACGGC 1617  
 Qy 243 CTGGGGATACACTCTGAGTCCCTCTGAG 271  
 Db 1618 CTGGGGATACACTCTGAGTCCCTCTGAG 1646





peptide nucleic acid-oligonucleotide (PNA) of at least 9 nucleotides and may form part of a set of probes for detecting the cytosine methylation state and/or single nucleotide polymorphisms and also to be used in an array for analysing diseases associated with CpG dinucleotides e.g. cancers and tumours. The probes can also be used in a method for ascertaining genetic and/or epigenetic parameters for the diagnosis and/or therapy of existing diseases or the predisposition to specific diseases, by analysing cytosine methylations. The parameters may be compared to another set of genetic and/or epigenetic parameters, the differences serving as basis for diagnosis and/or prognosis events which are disadvantageous to patients. The present sequence is one of the 533 genomic sequences derived from tumour suppressor genes and oncogenes. Sequences with even numbered Seq ID numbers are the complementary sequence of the corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence is missing).

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct](http://ftp.wipo.int/pub/published/pct) sequences.

CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published pct sequences.

SQ Sequence 6083 BP; 1525 A; 83 C; 1367 G; 3108 T; 0 other;

Query Match 60.6%; Score 164.2; DB 22; Length 6083;  
Best Local Similarity 76.2%; Pred. No. 4.4e-42;  
Matches 202; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY	6	CAAAATGCTGCTGAAATGCTTTTGGCATCTACGCTACTGACAGTAAAGCTTTCCCAATGCTC	65
Db	1324	CAAAATCACTTAAATATATATTTTAACTCACTAACTAAAGTAAATTTCCCAATGCTC	1365
QY	66	AACCTGTCCTGCGACGCTGATGAGGGGAGGAAAGGATTAACCTAGGGGTATGGGGGACCC	1325
Db	1264	AACCTATCTCTACCACTAATAAAAAATAAAAAAATAATTAACCTAAAAATATAAACGACC	1305
QY	126	AATCTGATGTCACCAACACTGACACAGCGGCATCCCGAGCTTTGCTGCTCACTACGCCCAA	185
Db	1204	AATCTTAAATCAACCACTAAACACGCGCCATCCCCAACTTATACCTCACTACGCCCAA	1145
QY	186	CTCCCAAGGGAGCAGCTATTTAAGGGGAGCAGAGGTGCAGAAACAAACAGAGGGCTG	245
Db	1144	CTCCCAAAAAAACAACCTATTTAATAAAAAAACAATAATACAAAAACAACAAACGACCTA	1085
QY	246	GGGATACAACTCTGGAGTCCTCTGA	270
Db	1084	AAAAATCAACTCTAAAAATCCTCTAA	1060

RESULT 9  
ABL70441  
ID ABL70441 standard; DNA; 5356 BP.  
XX  
AC ABL70441:

XX	01-JUL-2002 (first entry)
DT	
XX	
XX	Chemically treated cell signalling DNA sequence#166.
XX	
XX	Cell signalling; cytosine methylation; cell signalling disease;
KW	cancer; tumour; cytostatic; ds.
XX	
XX	Unidentified.
OS	

Olek A, Piepenbrock C, Berlin K;  
WPI; 2002-154758/20.

Nucleic acid, useful for diagnosis and therapy of diseases associated with cell signalling e.g. cancer, comprises chemically modified genomic sequences of genes associated with cell signalling -

Claim 1; SEQ ID NO 331; 24bp+sequence listing; English.

The invention relates to a nucleic acid comprising a sequence of at least 18 bases of a segment of chemically pretreated DNA of genes associated with cell signalling. The activity of the modified sequences of the invention may be described as cytostatic. The object of the invention is to provide the chemically modified DNA of genes associated with cell signalling, as well as oligonucleotides and/or PNA-oligomers for detecting cytosine methylations; as well as a method which is particularly suitable for the diagnosis and/or therapy of genetic and epigenetic parameters of genes associated with cell signalling. The chemically modified DNA provided by the invention is useful for diagnosis and therapy of diseases such as solid tumours and cancer. The sequences given in records AB170111-AB170626 represent chemically pre-treated genomic DNA's of genes associated with cell signalling.

Note: The sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

```

SQ Sequence 5356 BP; 1476 A; 60 C; 1356 G; 2464 T; 0 other;

Query Match      55.4%;   Score 150.2;   DB 24;   Length 5356;
Best Local Similarity 72.7%;   pred. No. 1.2e-37;
Matches 194;   Conservative 0;   Mismatches 73;   Indels 0;   Gaps 0;

```

QY	5	GCAAACTCGCCCTGAAATGTGTTTGGCATCACTACTGACACGTAAAGGTTTCCCAATCCT	64
Db	4032	GTAAATTTCTTTGAAATGTGTTTGGTATTAGTTATTGATACGTAAAGTTTTTTAAATTTT	4091
QY	65	CAACTCTGTCCTCCAGCTGATAGGGGAAGAAAGGATTACCTAGGGGTATGGGGCAC	124
Db	4092	TAAATTTTGTTTGTAGTTGATAGGGGAAGAAAGGATTATTAGGGGTATGGGGCAT	4151
QY	125	CAATCCTGAGTCCACCAACTGACACAGCCCATCCCCAGCCTTGTCCTCCTACCCCCA	184
Db	4152	TAAATTTGAGTTTATTAAATGATTACGTTATTTTATTAGTTTGTATTTATTATTTTA	4211
QY	185	ACCTCCACAGAGGACGCTATTAAAGGGGACGAGGTGCAGAACAAACAGACGSCCT	244
Db	4212	ATTTTGTAGGGAGTAGTTATTATTAGGGGAGTAGGAGTGTAGAATTAATAAGACGGTTT	4271
QY	245	GGGGATACAACTCTGGAGTCTCTGAG	271
Db	4272	GGGGATATAATTTTGGAGTTTTTTTGG	4298

RESULT 10	
ABK311472	
ID	ABK311472 standard; DNA; 5728 BP.
XX	
XX	
AC	ABK311472;
XX	
XX	
DT	23-APR-2002 (first entry)
XX	
XX	
DE	Signal transduction associated gene modified DNA #158.
XX	
XX	Human; signal transduction associated gene; cytosine methylation state;
KW	CpG island; signal transduction associated disease; solid tumour; cancer;
KW	antitumour; cytostatic; mutant; ds.
XX	
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
PN	WC200200926-A2.



QY 125 CAATCCTGAGTCACCACTGACACGCCCATCCCGAGCCTTGTGCTCACTACCCCA 184  
 DB 4879 TAATTTTGAGTTTATTAAATGATTACGTTTATTTTGTGTTTATTTTATTTA 4938  
 QY 185 ACCTCCAGAGGGGAGCTATTATTTAGGGGAGCAGAGTGCAGACAAACAGAGCGGCT 244  
 DB 4939 ATTTTGTAGAGGGAGTATTTTAAAGGGAGTAGGAGTGTAGATAAATAAGCGGTTT 4998  
 QY 245 GGGATACAACTCTGGAGTCTCTGAG 271  
 DB 4999 GGGATATAATTTTGGAGTTTGTGAG 5025

## RESULT 12

ABV09174/c

ID ABV09174 standard; cDNA; 204 BP.

XX AC ABV09174;

DT 13-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 9165.

DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 KW Homo sapiens.

OS WO200160860-A2.

XX PN 23-AUG-2001.

XX PD 20-FEB-2001; 2001WO-US05171.

XX PF 17-FEB-2000; 2000US-183319P.

XX PR 16-MAR-2000; 2000US-189862P.

XX PR 25-MAY-2000; 2000US-207454P.

XX PR 09-JUN-2000; 2000US-211314P.

XX PR 18-JUL-2000; 2000US-219007P.

XX PR 13-DEC-2000; 2000US-255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX PI WPI; 2001-662795/76.

XX DR Novel isolated nucleic acid molecule associated with cancerous state of

XX PT prostate cells and correlating with presence of prostate cancer, useful

XX PT for detecting presence of prostate cancer, stage of prostate cancer -

XX PS Claim 1; Page 1442; 11750pp; English.

XX CC The invention relates to an isolated nucleic acid molecule (I) comprising

XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

XX CC specification or its complement. (I) is useful for:

XX CC (a) assessing whether a patient is afflicted with prostate cancer;

XX CC (b) monitoring the progression of prostate cancer in a patient;

XX CC (c) assessing the efficacy of a test compound to inhibit prostate

XX CC cancer in a patient;

XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

XX CC in a patient;

XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;

XX CC (f) assessing the prostate cell carcinogenic potential of a compound;

XX CC (g) determining whether prostate cancer has metastasized in a patient;

XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a

XX CC patient;

XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX CC Sequence 204 BP; 26 A; 62 C; 55 G; 61 T; 0 other;

XX SQ Query Match 43.8%; Score 118.8; DB 23; Length 204;

Best Local Similarity 83.3%; Pred. No. 4.3e-28;  
 Matches 135; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 110 AGGGGTATGGGGACCAATCTCTGAGTCCACCAACTGACACGCCCATCCCGAGCCTTGTG 169

DB 191 AGGAAAAGAGCAACAGATCCAGGGAGCATTCACCTGCTCTCCAAACAGCCTTGTG 132

QY 170 CTTCACTACCTCCCACTCCAGAGGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGAA 229

DB 131 CTTCACTACCTCCCACTCCAGAGGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGAA 72

QY 230 CAAACAAGAGCGCTGGGATACAACTCTGGAGTCTCTCTGAG 271

DB 71 CAAACAAGAGCGCTGGGATACAACTCTGGAGTCTCTCTGAG 30

## RESULT 13

ABV39325

ID ABV39325 standard; cDNA; 481 BP.

XX AC ABV39325;

DT 16-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 39316.

DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 KW Homo sapiens.

OS WO200160860-A2.

XX PN 23-AUG-2001.

XX PD 20-FEB-2001; 2001WO-US05171.

XX PF 17-FEB-2000; 2000US-183319P.

XX PR 16-MAR-2000; 2000US-189862P.

XX PR 25-MAY-2000; 2000US-207454P.

XX PR 09-JUN-2000; 2000US-211314P.

XX PR 18-JUL-2000; 2000US-219007P.

XX PR 13-DEC-2000; 2000US-255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX PI WPI; 2001-662795/76.

XX DR Novel isolated nucleic acid molecule associated with cancerous state of

XX PT prostate cells and correlating with presence of prostate cancer, useful

XX PT for detecting presence of prostate cancer, stage of prostate cancer -

XX PS Claim 1; Page 7980; 11750pp; English.

XX CC The invention relates to an isolated nucleic acid molecule (I) comprising

XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

XX CC specification or its complement. (I) is useful for:

XX CC (a) assessing whether a patient is afflicted with prostate cancer;

XX CC (b) monitoring the progression of prostate cancer in a patient;

XX CC (c) assessing the efficacy of a test compound to inhibit prostate

XX CC cancer in a patient;

XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

XX CC in a patient;

XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;

XX CC (f) assessing the prostate cell carcinogenic potential of a compound;

XX CC (g) determining whether prostate cancer has metastasized in a patient;

XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a

XX CC patient;

XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX CC Sequence 481 BP; 120 A; 126 C; 142 G; 93 T; 0 other;

XX SQ



Query Match 43.8%; Score 118.8; DB 23; Length 481;  
 Best Local Similarity 83.3%; Pred. No. 5.7e-28;  
 Matches 135; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

2Y 110 AGGGTATGGCGACCAATCTGTAGTCCACCACTGACACGCCCTATCCCGACCTTTGTG 169  
 Db 60 AGGAAAAGAGCAACAGATCCAGGGAGCAATTCACCTGCCCTGTCTCCAAACAGCCTTTGTG 119

2Y 170 CCTCACTACCCCACTCCACAGGGAGCAGCTATTAAAGGGAGCAGAGTGCAGAA 229  
 Db 120 CCTCACTACCCCACTCCACAGGGAGCAGCTATTAAAGGGAGCAGAGTGCAGAA 179

2Y 230 CAACAAGACGGCCTGGGATACAACTCTGAGTCCTCTGAG 271  
 Db 180 CAACAAGACGGCCTGGGATACAACTCTGAGTCCTCTGAG 221

RESULT 14  
 ABK63588  
 ID ABK63588 standard; cDNA; 3330 BP.  
 AC ABK63588;  
 DE 18-JUN-2002 (first entry)

Rat sequence differentially expressed in response to a hepatotoxin #1495.  
 Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
 differential expression; centrilobular necrosis; steatosis.  
 Rattus norvegicus.  
 WO200210453-A2.  
 07-FEB-2002.  
 30-JUL-2001; 2001WO-US23872.  
 31-JUL-2000; 2000US-222040P.  
 02-NOV-2000; 2000US-244880P.  
 11-MAY-2001; 2001US-290029P.  
 15-MAY-2001; 2001US-290645P.  
 22-MAY-2001; 2001US-292336P.  
 06-JUN-2001; 2001US-295798P.  
 13-JUN-2001; 2001US-297457P.  
 19-JUN-2001; 2001US-298884P.  
 09-JUL-2001; 2001US-303459P.

(GENE-) GENE LOGIC INC.

Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

WPI; 2002-241625/29.

Predicting toxic effects of compounds or the progression of these toxic effects by determining the changes in gene expression in tissues or cells exposed to the toxin and comparing these to gene expression in unexposed tissues or cells -

Claim 1; Seq ID No 1495; 239pp; English.

The invention relates to methods for predicting toxic effects of compounds or the progression of these toxic effects by determining the global changes in gene expression in tissues or cells exposed to the toxin and comparing these to gene expression in unexposed tissues or cells. Also included are methods of predicting at least one toxic effect of a compound or progression of a toxic effect, preferably the hepatotoxicity of a compound, comprising detecting the level of expression in a tissue or cell sample exposed to the compound of two or more genes listed in the specification, where differential expression of the genes is indicative of at least one toxic effect or progression. The method can also be used to identify an agent which modulates the

toxic response and predict cellular pathways that a compound modulates in a cell. The methods utilise a set of at least two probes (on a solid support in kit form), where each of the probes comprises a sequence that specifically hybridises to a gene listed in the specification, a computer system comprising a database containing information identifying the expression level in a tissue or cell sample exposed to a hepatotoxin of a set of genes comprising at least two genes listed in the specification, and a user interface to view the information used to present information identifying the expression level in a tissue or cell of at least one gene listed in the specification. The method is useful for elucidating global changes in gene expression and for identifying toxicity markers in tissues or cell exposed to a known toxin. The genes may be used as toxicity markers in drug screening and toxicity assays. The genes and gene expression information may be used as diagnostic markers for the prediction or identification of the physiological state of tissue or cell sample that has been exposed to a compound or agent. Hepatotoxicity is characterised by centrilobular necrosis and steatosis. The present sequence is an expressed sequence tag (EST) or cDNA derived from a gene which is differentially expressed in response to a hepatotoxic agent.

SQ Sequence 3330 BP; 858 A; 860 C; 858 G; 753 T; 1 other;

Query Match 31.1%; Score 84.4; DB 24; Length 3330;

Best Local Similarity 64.9%; Pred. No. 1.1e-16;

Matches 157; Conservative 0; Mismatches 81; Indels 4; Gaps 2;

QY 6 CAATACTGCTGAAATGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAATCCTC 65  
 Db 260 CGAATCAGCTAAAGTTTATGATGGCCACCAACCATGGTATGAGGCTTTCCGGCCCTC 319

QY 66 AACTCTGTCTGCCAGCTGATGAGGGGAAGGAAGGATTTACCTAGGGGTATGG--CGA 123  
 Db 320 AAGCTGTCTGCCAGCTGTTGGGGGAAGGGAAATTTACCCAGGGGCTTGGGTATGC 379

QY 124 CCAATCTGAGTCCACCACTGACACGCCCA--TCCCAGCCTTGTGCTCACCACCC 181  
 Db 380 CCGTCTGTGAATCCATTATTGTGGCCACACCCACCTCCCATCTGTGGCTCTCCGATCC 439

QY 182 CCACCTCCCGAGGGAGCAGCTATTAAAGGGAGGAGGAGTGCAGAAACAAACAGACGG 241  
 Db 440 CCAGCCTGCGAGGGAGAGAGCTATTTAAGAGCATTTGGGAGTACAGGMAAACAGGCAG 499

QY 242 CC 243  
 Db 500 GC 501

RESULT 15

AAF15635 standard; cDNA; 1076 BP.

AC AAF15635;

XX AAF15635;

DT 13-MAR-2001 (first entry)

DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:70.

Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
 neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
 vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
 antibacterial; gene therapy; neural; immune; reproductive; renal;  
 gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
 wound; infectious disease; ss.

OS Homo sapiens.

XX WO200055174-A1.

PN 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX 12-MAR-1999; 99US-0124270.

```

XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
XX PI Rosen CA, Ruben SM;
XX
XX WPI; 2000-587513/55.
DR P-PSDB; AAB56432.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
XX Claim 1; Page 663; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
XX Sequence 1076 BP; 303 A; 281 C; 267 G; 220 T; 5 other;
SQ
Query Match 29.8%; Score 80.8; DB 21; Length 1076;
Best Local Similarity 95.3%; Pred. No. 1e-15;
Matches 82; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 186 CCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGAGTGCGAGCAACAAACAGCGCCTG 245
Db 2 CCAAACAGAGGAGCAGCTATTTAAGGGGAGCAGAGTGCGAGCAACAAACAGCGCCTG 61
Qy 246 GGGATACAACTCTGGAGTCCTCTGAG 271
Db 62 GGGATACAACTCTGGAGTCCTCTGAG 87

```

Search completed: February 19, 2003, 21:23:18  
Job time : 177.707 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

WM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 20:28:04 ; Search time 14.6331 Seconds  
(without alignments)  
2399.710 Million cell updates/sec

Title: US-09-808-388-5  
Perfect score: 271  
Sequence: 1 cggcgaaactgcctgaaa.....caactctggagctcctcgag 271

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Sequences: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/3A COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/3B COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30.6	11.3	305	4	US-09-328-111-618 Sequence 618, Appl
2	30.4	11.2	1549	2	US-08-856-444-1 Sequence 1, Appli
3	29.6	10.9	9299	3	US-08-458-434A-7 Sequence 7, Appli
4	29.2	10.8	3111	2	US-09-014-969-12 Sequence 12, Appl
5	29	10.7	204	4	US-09-506-729-37 Sequence 37, Appl
6	29	10.7	1883	1	US-08-202-056-2 Sequence 2, Appli
7	29	10.7	1933	1	US-08-076-093A-1 Sequence 1, Appli
8	29	10.7	1933	1	US-08-410-451-1 Sequence 1, Appli
9	29	10.7	1933	1	US-08-410-455-1 Sequence 1, Appli
10	29	10.7	1933	1	US-08-418-919-1 Sequence 1, Appli
11	29	10.7	1933	1	US-08-410-453A-2 Sequence 2, Appli
12	29	10.7	1933	1	US-08-701-265-1 Sequence 1, Appli
13	29	10.7	1933	1	US-08-410-454A-2 Sequence 2, Appli
14	29	10.7	1933	2	US-08-284-586-1 Sequence 1, Appli
15	29	10.7	1933	2	US-08-410-456A-2 Sequence 2, Appli
16	29	10.7	1933	2	US-08-805-478-1 Sequence 1, Appli
17	29	10.7	1933	2	US-08-802-627A-1 Sequence 1, Appli
18	29	10.7	1933	2	US-08-801-238-1 Sequence 1, Appli
19	29	10.7	1933	2	US-08-801-228-1 Sequence 1, Appli
20	29	10.7	1933	3	US-09-104-296-1 Sequence 1, Appli
21	29	10.7	1933	5	FCT-US94-06380-1 Sequence 1, Appli
22	28.6	10.6	3728	1	US-08-111-939-1 Sequence 1, Appli
23	28.4	10.5	2992	4	US-09-362-123A-3 Sequence 3, Appli
24	28.2	10.4	33	1	US-08-186-895-4 Sequence 4, Appli
25	28.2	10.4	4258	3	US-07-755-830A-5 Sequence 5, Appli
26	28	10.3	997	4	US-09-057-860A-3 Sequence 3, Appli
27	28	10.3	3100	1	US-08-296-362-1 Sequence 1, Appli

C	28	27.6	10.2	6803	3	US-08-665-259-19	Sequence 19, Appl
C	29	27.6	10.2	6803	3	US-08-762-500-19	Sequence 19, Appl
C	30	27.6	10.2	176373	3	US-09-128-155-17	Sequence 17, Appl
C	31	27	10.0	3021	4	US-09-556-877-182	Sequence 182, App
C	32	27	10.0	3021	4	US-09-620-412C-182	Sequence 182, App
C	33	27	10.0	3335	4	US-09-060-482-1	Sequence 1, Appli
C	34	27	10.0	7898	4	US-08-984-709A-49	Sequence 49, Appl
C	35	27	10.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C	36	26.8	9.9	1001	3	US-09-188-930-218	Sequence 218, App
C	37	26.8	9.9	1015	3	US-09-188-930-30	Sequence 30, Appl
C	38	26.6	9.8	238	4	US-08-903-233-128	Sequence 128, App
C	39	26.6	9.8	1452	2	US-08-770-544-7	Sequence 7, Appli
C	40	26.6	9.8	2458	3	US-09-071-101-5	Sequence 5, Appli
C	41	26.6	9.8	2458	3	US-09-369-618-6	Sequence 6, Appli
C	42	26.6	9.8	2458	3	US-09-369-617-6	Sequence 6, Appli
C	43	26.4	9.7	36519	3	US-08-923-137-2	Sequence 2, Appli
C	44	26.2	9.7	1356	2	US-08-484-126-4	Sequence 4, Appli
C	45	26	9.6	1440	3	US-08-652-285-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1  
US-09-328-111-618/c  
; Sequence 618, Application US/09328111  
; Patent No. 6262333  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Ascle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/328,111  
; CURRENT FILING DATE: 1999-06-08  
; EARLIER APPLICATION NUMBER: US 60/088,801  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 618  
; LENGTH: 305  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-328-111-618  
Query Match 11.3%; Score 30.6; DB 4; Length 305;  
Best Local Similarity 56.4%; Pred. No. 0.51;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
Qy 3 CGGCAAACTGCTGAAATGTTTGGCATCAGCTACTGACACGTAAAGTTTCCCAATC 62  
Db 217 CGCCAAATTAACCAAGCATGTTGTGTAACATCCCCCAGTGGGGGCTAGAAATTCCTCCATG 158  
Qy 63 CTCACCTCTGTCTGCCAGCTGATGAGGGGAAGGAAGGGA 103  
Db 157 GTGACCTGTGACCTGCTCCCTGAGACAGGGGAGCCAGGCA 117  
RESULT 2  
US-08-856-444-1  
; Sequence 1, Application US/08856444  
; Patent No. 5959081  
; GENERAL INFORMATION:

APPLICANT: Lecka-Czernik, Beata  
TITLE OF INVENTION: No. 5959081e1 Zinc Binding LIM Protein S2-6  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Benjamin Aaron Adler, Ph.D. J.D.  
STREET: 8011 Candle Lane  
CITY: Houston  
STATE: Texas  
ZIP: 77071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 Mb floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word for Macintosh  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,444  
FILING DATE: May 14, 1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benjamin Aaron Adler, Ph.D.  
REGISTRATION NUMBER: 35,423  
REFERENCE/DOCKET NUMBER: D5988  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (713) 777-2321  
TELEFAX: (713) 777-6908  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1549 bp  
TYPE: nucleic acid  
STRANDEDNESS: single-stranded  
TOPOLOGY: linear  
MOLECULE TYPE: c-DNA  
DESCRIPTION: NO  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
IMMEDIATE SOURCE:  
POSITION IN GENOME:  
FEATURE:  
PUBLICATION INFORMATION:  
US-08-856-444-1  
Query Match 11.2%; Score 30.4; DB 2; Length 1549;  
Best Local Similarity 57.3%; Pred. No. 1.2;  
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
QY 50 AGGTTTCCCAATCCTCAACTGTCTCTCCAGCTGATGAGGGGAAGGAAAGGATTACT 109  
DB 530 AAGTGGCCCTCCGGCAGGTGCTTGGCCAGGAGGAGGGAAGCAGCAGGAAAGCC 589  
QY 110 AGGGGTATGGCGACCAATCTGAGTCCACCACTG 145  
DB 590 AGAGGGGCGAGACCACTGCTGTACCACCAACGG 625  
RESULT 3  
US-08-458-434A-7/C  
Sequence 7, Application US/08458434A  
Patent No. 6083690  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Stephen E.  
APPLICANT: Mundy M.D., Gregory R.  
APPLICANT: Gosh-Choudhury Ph.D., Nandini  
APPLICANT: Peng Ph.D., Jian Q.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING  
TITLE OF INVENTION: OSTEOGENIC AGENTS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: James C. Weseman, Esq.

STREET: 401 B. Street, Suite 1700  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,434A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Weseman, James C.  
REGISTRATION NUMBER: 30,507  
REFERENCE/DOCKET NUMBER: P0060U0S0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 699-3604  
TELEFAX: 619-236-1048  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9299 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-458-434A-7  
Query Match 10.9%; Score 29.6; DB 3; Length 9299;  
Best Local Similarity 59.5%; Pred. No. 4.6;  
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
QY 150 CGCCCATCCCGACCTTGCTCACTACCCCAACCTCCAGAGGAGGAGCTATTATTA 209  
DB 6318 CCCCCACCCCGCCCTTCTCTCGCCCTCCAGCCCAATTCACAACTTCAGCTGGTTA 6259  
QY 210 AGGGGAGCAGGAGTGCGAACAA 233  
DB 6258 AGAACAGGAGGAGGGGAGAACAGA 6235  
RESULT 4  
US-09-014-969-12  
Sequence 12, Application US/09014969  
Patent No. 5965397  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John W.  
APPLICANT: Lavallie, Edward R.  
APPLICANT: Racie, Lisa A.  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/014,969



```

;
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,093A
; FILING DATE: 11-Jun-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706F2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1933 nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-076-093A-1

Query Match 10.7%; Score 29; DB 1; Length 1933;
Best Local Similarity 57.0%; Pred. No. 3.8;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 4 GGCATAATGCTGCTGAAATGTTTGGCATCAGCTACTGACAGTAAAGTTTCCCAATCC 63
DB 1915 GGAACATCTGCTGCCAATGGACTGGTGGCTGCACATGCTTCTAGGGATGCTGATGC 1856

QY 64 TCACTCTGCTGCCAGCTGATGAGGGGAGG 96
DB 1855 TGCACGCCAGCTGGAAGCTGCAGAGGGGAGG 1823

RESULT 8
US-08-410-451-1/c
; Sequence 1, Application US/08410451
; Patent No. 5552284
; GENERAL INFORMATION:
; APPLICANT: Lee, James,
; APPLICANT: Holmes, William E.,
; APPLICANT: Woods, William I.
; TITLE OF INVENTION: Human PF4A Receptors and Their Use
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,451
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/234494
; FILING DATE: 28-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706C1DS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
;

```

```

;
; FILING DATE: 28-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706C1D4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1933 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-410-451-1

Query Match 10.7%; Score 29; DB 1; Length 1933;
Best Local Similarity 57.0%; Pred. No. 3.8;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 4 GGCATAATGCTGCTGAAATGTTTGGCATCAGCTACTGACAGTAAAGTTTCCCAATCC 63
DB 1915 GGAACATCTGCTGCCAATGGACTGGTGGCTGCACATGCTTCTAGGGATGCTGATGC 1856

QY 64 TCACTCTGCTGCCAGCTGATGAGGGGAGG 96
DB 1855 TGCACGCCAGCTGGAAGCTGCAGAGGGGAGG 1823

RESULT 9
US-08-410-455-1/c
; Sequence 1, Application US/08410455
; Patent No. 5571702
; GENERAL INFORMATION:
; APPLICANT: Lee, James,
; APPLICANT: Holmes, William E.,
; APPLICANT: Woods, William I.
; TITLE OF INVENTION: Human PF4A Receptors and Their Use
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,455
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/234494
; FILING DATE: 28-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706C1DS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
;

```

INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1933 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
JS-08-410-455-1

Query Match 10.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.8;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

2Y 4 GGCAAACTGCTGAAATGTTTGGCATCAGTACTACACGTAAAGTTTCCCAATCC 63  
2b 1915 GGAACATCTGCTGCCAATGAGTGTGGTGCATGCTTCTTAGGGATGCTGATGC 1856

2Y 64 TCAACTCTGCTGCCACCTCATGAGGGGAAGG 96

2b 1855 TGCACGCCAGCTGGAAGCTGCAGAGGGGAAGG 1823

## RESULT 10

JS-08-418-919-1/c  
; Sequence 1, Application US/08418919  
; Patent No. 5633141

; GENERAL INFORMATION:  
; APPLICANT: Lee, James,  
; APPLICANT: Holmes, William E.,  
; APPLICANT: Woods, William I.  
; TITLE OF INVENTION: Human PFAA Receptors and Their Use  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080

## COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/418,919

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/234,494

FILING DATE:

APPLICATION NUMBER: 07/677,211

FILING DATE: 29-Mar-1991

ATTORNEY/AGENT INFORMATION:

NAME: Hensley, Max D.

REGISTRATION NUMBER: 27,043

REFERENCE/DOCKET NUMBER: 706

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/266-1994

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1933 bases

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

JS-08-418-919-1

Query Match 10.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.8;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

2Y 4 GGCAAACTGCTGAAATGTTTGGCATCAGTACTACACGTAAAGTTTCCCAATCC 63

Db 1915 GGAACATCTGCTGCCAATGAGTGTGGTGCATGCTTCTTAGGGATGCTGATGC 1856  
QY 64 TCAACTCTGCTGCCAGCTGATGAGGGGAAGG 96  
Db 1855 TGCACGCCAGCTGGAAGCTGCAGAGGGGAAGG 1823

## RESULT 11

US-08-410-453A-2/c  
; Sequence 2, Application US/08410453A  
; Patent No. 5767063

; GENERAL INFORMATION:  
; APPLICANT: Lee, James,  
; APPLICANT: Holmes, William E.,  
; APPLICANT: Woods, William I.  
; TITLE OF INVENTION: Human PFAA Receptors and Their Use  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/410,453A

FILING DATE: 24-Mar-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/234494

FILING DATE: 28-APR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/677211

FILING DATE: 29-Mar-1991

ATTORNEY/AGENT INFORMATION:

NAME: Love, Richard B.

REGISTRATION NUMBER: 34,659

REFERENCE/DOCKET NUMBER: P0706C1D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-5530

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1933 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-08-410-453A-2

Query Match 10.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.8;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 4 GGCAAACTGCTGAAATGTTTGGCATCAGTACTACACGTAAAGTTTCCCAATCC 63

Db 1915 GGAACATCTGCTGCCAATGAGTGTGGTGCATGCTTCTTAGGGATGCTGATGC 1856

QY 64 TCAACTCTGCTGCCAGCTGATGAGGGGAAGG 96

Db 1855 TGCACGCCAGCTGGAAGCTGCAGAGGGGAAGG 1823

## RESULT 12

US-08-701-265-1/c

; Sequence 1, Application US/08701265

; Patent No. 5776457

; GENERAL INFORMATION:

APPLICANT: Chuntharapai, Anan  
APPLICANT: Lee, James  
APPLICANT: Hebert, Caroline  
APPLICANT: Jin Kim, K.  
TITLE OF INVENTION: Antibodies to Human PF4A Receptors  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,265  
FILING DATE: 22-AUG-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/076093  
FILING DATE: 11-Jun-1993  
APPLICATION NUMBER: 07/810782  
FILING DATE: 19-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/677211  
FILING DATE: 29-MAR-1991  
NAME: Love, Richard B.  
REGISTRATION NUMBER: 34,659  
REFERENCE/DOCKET NUMBER: 706P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/952-9881  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 nucleotides  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-701-265-1

Query Match 10.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.8;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 4 GGCMAACTGCTGAAATGCTTTGGCATCAGCTACTGACACGTAAGGTTCCCAATCC 63  
DB 1915 GGAACATCTGCTGCCCAATGAGTGGCTGACATGGCTTTCTAGGATGCTGATGC 1856

QY 64 TCAACTCTGCTGCCAGCTGATGAGGGGAAGG 96  
DB 1855 TGCACCCAGCTGGAAGTGCAGAGGGGAAGG 1823

RESULT 13  
US-08-410-454A-2/c  
Sequence 2, Application US/08410454A  
Patent No. 5783415  
GENERAL INFORMATION:  
APPLICANT: Lee, James  
APPLICANT: Holmes, William E.,  
APPLICANT: Woods, William I.  
TITLE OF INVENTION: Human PF4A Receptors and Their Use  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco

STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/410,454A  
FILING DATE: 24-Mar-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/234494  
FILING DATE: 28-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/677211  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B.  
REGISTRATION NUMBER: 34,659  
REFERENCE/DOCKET NUMBER: P0706C1D3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5530  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-410-454A-2

Query Match 10.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.8;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 4 GGCMAACTGCTGAAATGCTTTGGCATCAGCTACTGACACGTAAGGTTCCCAATCC 63  
DB 1915 GGAACATCTGCTGCCCAATGAGTGGCTGACATGGCTTTCTAGGATGCTGATGC 1856

QY 64 TCAACTCTGCTGCCAGCTGATGAGGGGAAGG 96  
DB 1855 TGCACCCAGCTGGAAGTGCAGAGGGGAAGG 1823

RESULT 14  
US-08-284-586-1/c  
Sequence 1, Application US/08284586  
Patent No. 5840856  
GENERAL INFORMATION:  
APPLICANT: Chuntharapai, Anan  
APPLICANT: Lee, James  
APPLICANT: Hebert, Caroline  
APPLICANT: Jin Kim, K.  
TITLE OF INVENTION: Antibodies to Human PF4A Receptors  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/284,586  
FILING DATE:



CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/076,093A  
FILING DATE: 11-Jun-1993  
APPLICATION NUMBER: 07/810782  
FILING DATE: 19-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/677211  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B  
REGISTRATION NUMBER: 34,659  
REFERENCE/DOCKET NUMBER: 706P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5530  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 nucleotides  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
S-08-284-586-1

Query Match 10.7%; Score 29; DB 2; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.8;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

y 4 GGCAAACTGCTGAAATGTTTGGCATCAGTCTGACACGTAAGGTTTCCCAATCC 63  
b 1915 GGAACATCTGCTGCCCAATGGACTGTGGCTGCACATGGCTTTCTAGGGATGCTGATGC 1856  
y 64 TCAACTCTGCTGCCAGCTGATGAGGGGAAGG 96  
b 1855 TGCACGCCAGCTGGAGCTGCAGAGGGGAAGG 1823

RESULT 15  
S-08-410-456A-2/c  
Sequence 2, Application US/08410456A  
Patent No. 5856457  
GENERAL INFORMATION:  
APPLICANT: Lee, James  
APPLICANT: Holmes, William E.,  
APPLICANT: Woods, William I.  
TITLE OF INVENTION: Human PF4A Receptors and Their Use  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/410,456A  
FILING DATE: 24-Mar-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/234494  
FILING DATE: 28-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/677211  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B.  
REGISTRATION NUMBER: 34,659

REFERENCE/DOCKET NUMBER: P0706C1D2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-5530  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-410-456A-2

Query Match 10.7%; Score 29; DB 2; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.8;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
y 4 GGCAAACTGCTGAAATGTTTGGCATCAGTCTGACACGTAAGGTTTCCCAATCC 63  
b 1915 GGAACATCTGCTGCCCAATGGACTGTGGCTGCACATGGCTTTCTAGGGATGCTGATGC 1856  
y 64 TCAACTCTGCTGCCAGCTGATGAGGGGAAGG 96  
b 1855 TGCACGCCAGCTGGAGCTGCAGAGGGGAAGG 1823

Search completed: February 20, 2003, 01:16:27  
Job time : 42.6331 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

M nucleic - nucleic search, using sw model

un on: February 19, 2003, 23:17:14 ; Search time 205.564 Seconds  
(without alignments)  
671.453 Million cell updates/sec

file: US-09-808-388-5

effect score: 271

sequence: 1 cgcggcaaacgctgctgaa.....caactctggagctctctgag 271

coring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

searched: 424239 seqs, 254661826 residues

total number of hits satisfying chosen parameters: 848478

inimum DB seq length: 0

aximum DB seq length: 200000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/ECT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PTU5\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	length	ID	Description
1	271	100.0	271	10	US-09-808-388-5
2	271	100.0	332	10	US-09-808-388-6
3	84.4	31.1	3330	10	US-09-917-800A-1495
4	80.8	29.8	1076	10	US-09-925-300-70
5	50	18.5	735	9	US-09-981-333-17
6	33.8	12.5	2136	10	US-09-862-658-3
7	33.8	12.5	3320	10	US-09-862-658-1
8	31.4	11.6	771	9	US-10-010-920-4
9	31.4	11.6	771	9	US-10-010-920-5
10	31.4	11.6	955	9	US-10-010-920-3
11	30.6	11.3	305	10	US-09-878-536-618
12	30.6	11.3	2885	10	US-09-880-107-3388
13	30.6	11.3	3370	12	US-10-044-090-339
14	30.2	11.1	187	10	US-09-783-590-3051
15	30.2	11.1	340	10	US-09-833-381-1715
16	30.2	11.1	30676	10	US-09-927-091-8
17	30.2	11.1	45845	10	US-09-927-091-6
18	29.8	11.0	854	10	US-09-763-231A-42
19	29.2	10.8	41936	10	US-09-967-768A-116

20	29	10.7	544	9	US-09-854-133-474	Sequence 474, App
21	29	10.7	544	10	US-09-738-973-474	Sequence 474, App
22	29	10.7	1933	9	US-09-104-063-1	Sequence 1, Appli
23	28.8	10.6	510	10	US-09-783-590-9949	Sequence 9949, Ap
24	28.8	10.6	2223	9	US-09-984-245-120	Sequence 120, App
25	28.6	10.6	403	10	US-09-878-574-2747	Sequence 2747, Ap
26	28.6	10.6	5950	10	US-09-864-864-277	Sequence 277, App
27	28.6	10.6	49744	10	US-09-927-091-4	Sequence 4, Appli
28	28.4	10.5	4420	9	US-09-986-234-5	Sequence 5, Appli
29	28.4	10.5	7521	9	US-09-819-104A-3	Sequence 3, Appli
30	28.4	10.5	8686	9	US-09-819-104A-1	Sequence 1, Appli
31	28.2	10.4	397	10	US-09-960-352-2124	Sequence 2124, Ap
32	28.2	10.4	452	9	US-09-736-457-829	Sequence 829, App
33	28.2	10.4	452	9	US-09-736-457-1083	Sequence 1083, Ap
34	28.2	10.4	452	9	US-09-902-941-829	Sequence 829, App
35	28.2	10.4	452	9	US-09-902-941-1083	Sequence 1083, Ap
36	28.2	10.4	452	9	US-09-849-626-829	Sequence 829, App
37	28.2	10.4	452	9	US-09-849-626-1083	Sequence 1083, Ap
38	28.2	10.4	452	10	US-09-920-300A-676	Sequence 676, App
39	28.2	10.4	452	12	US-10-033-528-676	Sequence 676, App
40	28.2	10.4	1119	10	US-09-741-669-174	Sequence 174, App
41	28.2	10.4	1119	10	US-09-912-020-215	Sequence 215, App
42	28	10.3	468	10	US-09-864-761-5528	Sequence 5528, Ap
43	28	10.3	78056	9	US-10-109-551-1	Sequence 1, Appli
44	27.8	10.3	319	10	US-09-960-352-9985	Sequence 9985, Ap
45	27.8	10.3	493	10	US-09-783-590-6591	Sequence 6591, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-808-388-5  
; Sequence 5, Application US/09808388  
; Patent No. US20020081719A1  
; GENERAL INFORMATION:  
; APPLICANT: Massaad, Charbel  
; APPLICANT: Berenbaum, Francis  
; APPLICANT: Olivier, Jean-Luc  
; APPLICANT: Salvat, Colette  
; APPLICANT: Berezziat, Gilbert  
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
; FILE REFERENCE: ST00010  
; CURRENT APPLICATION NUMBER: US/09/808,388  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: FR/00/03262  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: US 60/196,959  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 5  
; LENGTH: 271  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fragment of the PLA2s promoter  
US-09-808-388-5

Query Match	100.0%	Score	271;	DB	10;	Length	271;
Best Local Similarity	100.0%	Pred. No.	1.1e-81;				
Matches	271;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	CGCGGCAAACTGCTTGAATGCTTTGGCATCAGTACTACACAGGTAAGGTTTCCCAA	60				
Db	1	CGCGGCAAACTGCTTGAATGCTTTGGCATCAGTACTACACAGGTAAGGTTTCCCAA	60				
QY	61	TCCTCAACTCTGTCTTCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGG	120				
Db	61	TCCTCAACTCTGTCTTCCAGCTGATGAGGGGAAGAAAGGATTACCTAGGGGTATGGG	120				
QY	121	CGACCAATCTGTAGTCCCACTGACACAGGCCCAATCCCGACCTTGCTGCCTCACCTACC	180				

Db 121 CGACCAATCCTGAGTCCACCAACTGACCAAGCCCATCCAGCCCTTGCGCTCACCTACC 180  
QY 181 CCCAACTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCAGAAACAAACAGACG 240  
Db 181 CCCAACTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCAGAAACAAACAGACG 240  
QY 241 GCGTGGGATACAACTCTCGAGTCTCTGAG 271  
Db 241 GCGTGGGATACAACTCTCGAGTCTCTGAG 271

RESULT 2

US-09-808-388-6  
; Sequence 6, Application US/09808388  
; Patent No. US20020081719A1  
; GENERAL INFORMATION:  
; APPLICANT: Massaad, Charbel  
; APPLICANT: Berenbaum, Francis  
; APPLICANT: Olivier, Jean-Luc  
; APPLICANT: Salvat, Colette  
; APPLICANT: Berziat, Gilbert  
; TITLE OF INVENTION: Inducible Hybrid Promoters, Vectors Comprising them  
; FILE REFERENCE: ST00010  
; CURRENT APPLICATION NUMBER: US/09/808,388  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: FR/00/03262  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: US 60/196,959  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 332  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: PPRE/PLA2s hybrid promoter  
US-09-808-388-6

Query Match 100.0%; Score 271; DB 10; Length 332;  
Best Local Similarity 100.0%; Pred. No. 1.28-81;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGCGGCAAACTGCCTGAAATGTTTGGCATCAGTACTGACACGTAAGGTTTCCCAA 60  
Db 62 CGCGGCAAACTGCCTGAAATGTTTGGCATCAGTACTGACACGTAAGGTTTCCCAA 121  
QY 61 TCCTCAACTCTGCTCCAGTCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGGGGTATGG 120  
Db 122 TCCTCAACTCTGCTCCAGTCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGGGGTATGG 181  
QY 121 CGACCAATCCTGAGTCCACCACTGACACGCGCCATCCCGAGCTTGTGCTCAGCTACC 180  
Db 182 CGACCAATCCTGAGTCCACCACTGACACGCGCCATCCCGAGCTTGTGCTCAGCTACC 241  
QY 181 CCCAACTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCAGAAACAAACAGACG 240  
Db 242 CCCAACTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCAGAAACAAACAGACG 301

QY 241 GCGTGGGATACAACTCTCGAGTCTCTGAG 271  
Db 302 GCGTGGGATACAACTCTCGAGTCTCTGAG 332

RESULT 3

US-09-917-800A-1495  
; Sequence 1495, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark

; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1495  
; LENGTH: 3330  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 X51529  
US-09-917-800A-1495

Query Match 31.1%; Score 84.4; DB 10; Length 3330;  
Best Local Similarity 64.9%; Pred. No. 1.9e-15;  
Matches 157; Conservative 0; Mismatches 81; Indels 4; Gaps 2;  
QY 6 CAAACTGCTGAAATGTTTGGCATCAGTACTGACACGTAAGGTTTCCCAATCCTC 65  
Db 260 CGAAATCAGCTAAAGTTATGATGCGCACACCCATGATAGGGGCTTTTCGGGCTC 319  
QY 66 AACTGTGCTCCTGCCAGTCTGAGGGGAGGAAAGGATTAACCTAGGGGTATGG--CGA 123  
Db 320 AAGGCTGTCTGCCAGTCTTGGGGGAAAGGGGAAATTAACCCAGGCGCTTGGGTATGC 379  
QY 124 CCAATCCTGAGTCCACCACTGACACGCGCCA--TCCCGAGCCTTGTGCTCAGCTACCC 181  
Db 380 CCGTCTGTGAATCCATTTATTTGGCCACACCCACCTCCCATCTCCCTGTGGCTCTCGATCC 439  
QY 182 CCAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTGCAGAAACAAACAGACGG 241  
Db 440 CCAGCCTCGACAGGAGAGAGCTATTTAAGAGCATTTGGAGTACAGGAAAAACAAGGACG 499  
QY 242 CC 243  
Db 500 GC 501

RESULT 4

US-09-925-300-70  
; Sequence 70, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben,  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCI/US00/05988

PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 70  
LENGTH: 1076  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (911)  
OTHER INFORMATION: n equals a,t,g, or c  
S-09-925-300-70

Query Match 29.8%; Score 80.8; DB 10; Length 1076;  
Best Local Similarity 95.3%; Pred. No. 1.9e-17;  
Matches 82; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
Y 186 CTTCCAGAGGAGCAGCTATTTAAGGGAGCAGGAGTGAGAAACAAACAGAGCGCCTG 245  
b 2 CCAACAGAGGAGCAGCTATTTAAGGGAGCAGGAGTGAGAAACAAACAGAGCGCCTG 61  
Y 246 GGGATACAACTCTGGAGTCTCTGAG 271  
b 62 GGGATACAACTCTGGAGTCTCTGAG 87

RESULT 5  
S-09-981-353-17  
Sequence 17, Application US/09981353  
Patent No. US20020160382A1  
GENERAL INFORMATION:  
APPLICANT: Lasek, Amy W.  
APPLICANT: Jones, David A.  
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
FILE REFERENCE: PA-0038 US  
CURRENT APPLICATION NUMBER: US/09/981,353  
CURRENT FILING DATE: 2001-10-11  
NUMBER OF SEQ ID NOS: 194  
SOFTWARE: PERL Program  
SEQ ID NO 17  
LENGTH: 735  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20020160382A1 474322.36  
NAME/KEY: unsure  
LOCATION: 388  
OTHER INFORMATION: a, t, c, g, or other  
IS-09-981-353-17

Query Match 18.5%; Score 50; DB 9; Length 735;  
Best Local Similarity 100.0%; Pred. No. 3.7e-07;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 222 GTGCGAACAACAGAGCGCCTGGGATACAACTCTGGAGTCTCTGAG 271  
b 1 GTGCGAACAACAAGAGCGCCTGGGATACAACTCTGGAGTCTCTGAG 50

RESULT 6  
IS-09-862-658-3  
Sequence 3, Application US/09862658  
Patent No. US20020137101A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
TITLE OF INVENTION: 46638, A NOVEL HUMAN LIPOXYGENASE FAMILY  
FILE REFERENCE: 10448-053001  
CURRENT APPLICATION NUMBER: US/09/862,658  
CURRENT FILING DATE: 2001-05-22

PRIOR APPLICATION NUMBER: US 60/205,675  
PRIOR FILING DATE: 2000-05-19  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 2136  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-862-658-3

Query Match 12.5%; Score 33.8; DB 10; Length 2136;  
Best Local Similarity 53.4%; Pred. No. 0.17;  
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 68 CTCTGTCTCTCCAGCTGATGAGGGAAGAAAGGATACCTAGGGGTATGGGACCAA 127  
Db 1752 CTCTGCCAGCAGCTGTCTCAAGTGGCAGCATGACTTTGGGCTGTGATGCCAA 1811  
QY 128 TCTGTAGTCCACCACTGACCAAGCCCATGCCAGCCTTGTGCCTACCTACCCCAACC 187  
Db 1812 TGTCTCATCATCATGAGCAGCCCCACCCAGACCAAGGGGACCAACCCCTGAAGAC 1871  
QY 188 TCCAGAGGGAGC 200  
Db 1872 TTACTAGACACC 1884

RESULT 7  
US-09-862-658-1  
Sequence 1, Application US/09862658  
Patent No. US20020137101A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
TITLE OF INVENTION: 46638, A NOVEL HUMAN LIPOXYGENASE FAMILY  
FILE REFERENCE: 10448-053001  
CURRENT APPLICATION NUMBER: US/09/862,658  
CURRENT FILING DATE: 2001-05-22  
PRIOR APPLICATION NUMBER: US 60/205,675  
PRIOR FILING DATE: 2000-05-19  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 3320  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (459)...(2591)  
US-09-862-658-1

Query Match 12.5%; Score 33.8; DB 10; Length 3320;  
Best Local Similarity 53.4%; Pred. No. 0.2;  
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 68 CTCTGTCTCTCCAGCTGATGAGGGAAGAAAGGATACCTAGGGGTATGGGACCAA 127  
Db 2210 CTCTGCCAGCAGCTGTCTCAAGTGGCAGCATGACTTTGGGCTGTGATGCCAA 2269  
QY 128 TCTGTAGTCCACCACTGACCAAGCCCATGCCAGCCTTGTGCCTACCTACCCCAACC 187  
Db 2270 TGTCTCATCATCATGAGCAGCCCCACCCAGACCAAGGGGACCAACCCCTGAAGAC 2329  
QY 188 TCCAGAGGGAGC 200  
Db 2330 TTACTAGACACC 2342

RESULT 8  
US-10-010-920-4/c  
Sequence 4, Application US/10010920  
Publication No. US20030027165A1  
GENERAL INFORMATION:

```
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Alternatively spliced polk nucleotide and amino acid sequences
; FILE REFERENCE: 98,723-E3
; CURRENT APPLICATION NUMBER: US/10/010,920
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,649
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-010-920-4

Query Match      11.6%; Score 31.4; DB 9; Length 771;
Best Local Similarity 57.7%; Pred. No. 0.69;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 175 CCTACCCCAACCTCCAGAGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGACAAAC 234
Db 578 CCTACCTCCGGCTCTCCCGGGTGACGACGGGTAGAAAAGCAGGAGCGGAGAAAGGAG 519

QY 235 AAGACGGCCTGGGATACAACTCTGGAGTCTCTGAG 271
Db 518 AGGCGGGGTAGGATGCAGCTGTGTCATTCTGGG 482

RESULT 9
US-10-010-920-5
; Sequence 5, Application US/10010920
; Publication No. US20030027165A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Alternatively spliced polk nucleotide and amino acid sequences
; FILE REFERENCE: 98,723-E3
; CURRENT APPLICATION NUMBER: US/10/010,920
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,649
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-010-920-5

Query Match      11.6%; Score 31.4; DB 9; Length 771;
Best Local Similarity 57.7%; Pred. No. 0.69;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 175 CCTACCCCAACCTCCAGAGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGACAAAC 234
Db 194 CCTACCTCCGGCTCTCCCGGGTGACGACGGGTAGAAAAGCAGGAGCGGAGAAAGGAG 253

QY 235 AAGACGGCCTGGGATACAACTCTGGAGTCTCTGAG 271
Db 254 AGGCGGGGTAGGATGCAGCTGTGTCATTCTGGG 290

RESULT 10
US-10-010-920-3/c
; Sequence 3, Application US/10010920
; Publication No. US20030027165A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Alternatively spliced polk nucleotide and amino acid sequences
; FILE REFERENCE: 98,723-E3
; CURRENT APPLICATION NUMBER: US/10/010,920
```

```
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,649
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 955
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-010-920-3

Query Match      11.6%; Score 31.4; DB 9; Length 955;
Best Local Similarity 57.7%; Pred. No. 0.75;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 175 CCTACCCCAACCTCCAGAGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGACAAAC 234
Db 727 CCTACCTCCGGCTCTCCCGGGTGACGACGGGTAGAAAAGCAGGAGCGGAGAAAGGAG 668

QY 235 AAGACGGCCTGGGATACAACTCTGGAGTCTCTGAG 271
Db 667 AGGCGGGGTAGGATGCAGCTGTGTCATTCTGGG 631

RESULT 11
US-09-879-536-618/c
; Sequence 618, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astie, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879,536
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/088,801
; PRIOR FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 618
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-879-536-618

Query Match      11.3%; Score 30.6; DB 10; Length 305;
Best Local Similarity 56.4%; Pred. No. 0.85;
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 3 CGGCAAAACCTGCCTGAAATGTGTTTGGCATCAGCTACTGACACAGTAAGGTTTCCCAATC 62
Db 217 CGCCAAAATAACAGCATGTGTGTGTAAATCCCCCAGTGGGGCTAGAAATTTCCCATG 158

QY 63 CTCAACTCTGTCTGCCAGCTGATGAGGGAAGGAAGGA 103
Db 157 GTGACCTGTGACCTGCTCCTCTGAGACAGGGAGGCCAGCA 117

RESULT 12
US-09-880-107-3398/c
; Sequence 3388, Application US/09880107
; Patent No. US20020142981A1
```

GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3388  
LENGTH: 2885  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 U53347  
S-09-880-107-3388  
Query Match 11.3%; Score 30.6; DB 10; Length 2885;  
Best Local Similarity 56.4%; Pred. No. 2.3;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
Y 3 CGGCAAACTGCTGAAATGTTTGGCATCAGTACTGACACGTAAGTTTCCCAATC 62  
b 2633 CGCCAAATAACCAAGCATGTTGTTAAATCCATCCCCAGTGGGGCTAGAAATCCCATG 2574  
Y 63 CTCACCTCTCTCTGCGACGCTGATAGGGGGAAGGGA 103  
b 2573 GTGACCTGTGACCTGCTCTCTGAGACAGGGAGGCCAGGCA 2533  
RESULT 13  
S-10-044-090-339/c  
Sequence 339, Application US/10044090  
Patent No. US20020137081A1  
GENERAL INFORMATION:  
APPLICANT: Olga Bandman  
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
FILE REFERENCE: PA-0028 US  
CURRENT APPLICATION NUMBER: US/10/044,090  
CURRENT FILING DATE: 2002-01-09  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: PERL Program  
SEQ ID NO 339  
LENGTH: 3370  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20020137081A1 346535.6  
IS-10-044-090-339  
Query Match 11.3%; Score 30.6; DB 12; Length 3370;  
Best Local Similarity 56.4%; Pred. No. 2.4;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
Y 3 CGGCAAACTGCTGAAATGTTTGGCATCAGTACTGACACGTAAGTTTCCCAATC 62  
b 3141 CGCCAAATAACCAAGCATGTTGTTAAATCCATCCCCAGTGGGGCTAGAAATCCCATG 3082  
Y 63 CTCACCTCTGCTCTGCGACGCTGATAGGGGGAAGGGA 103  
b 3081 GTGACCTGTGACCTGCTCTCTGAGACAGGGAGGCCAGGCA 3041  
RESULT 14  
IS-09-783-590-3051/c  
Sequence 3051, Application US/09783590

Patent No. US20020110850A1  
GENERAL INFORMATION:  
APPLICANT: Dillon, Patrick J.  
APPLICANT: Haseltine, William A.  
APPLICANT: Li, Haodong  
APPLICANT: Rosen, Craig A.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
FILE REFERENCE: PO-16.2C1  
CURRENT APPLICATION NUMBER: US/09/783,590  
CURRENT FILING DATE: 2000-02-15  
PRIOR APPLICATION NUMBER: 08/420,856  
PRIOR FILING DATE: 1995-04-12  
PRIOR APPLICATION NUMBER: 08/346,731  
PRIOR FILING DATE: 1994-11-21  
NUMBER OF SEQ ID NOS: 12485  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3051  
LENGTH: 187  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (10)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (19)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (157)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (159)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (164)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-3051  
Query Match 11.1%; Score 30.2; DB 10; Length 187;  
Best Local Similarity 58.2%; Pred. No. 0.93;  
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;  
QY 170 CCTCAGCTACCCCAACCTCCAGAGGGAGCAGCTATTTAAGGGGAGCAGAGTCCAGAA 229  
Db 133 CCACGGCAGCCAGGACACCCGCTAGGAGGAGCAGGTACCGAGGCGGGAGAGCGCA 74  
QY 230 CAACAAGACGGCTGGGATACAACTCTGG 260  
Db 73 CAACGGCAGACCTCGGACTCCACCCGG 43  
RESULT 15  
US-09-833-381-1715  
Sequence 1715, Application US/09833381  
Patent No. US20020132090A1  
GENERAL INFORMATION:  
APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs  
FILE REFERENCE: 5800-119  
CURRENT APPLICATION NUMBER: US/09/833,381  
CURRENT FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 09/516,448  
PRIOR FILING DATE: 2000-02-29  
NUMBER OF SEQ ID NOS: 2050  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1715  
LENGTH: 340  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-833-381-1715  
Query Match 11.1%; Score 30.2; DB 10; Length 340;



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

3M nucleic - nucleic search, using sw model

run on: February 19, 2003, 19:33:48 ; Search time 1318.61 Seconds

(without alignment)  
3328.484 Million cell updates/sec

Title: US-09-808-388-5

Perfect score: 271

Sequence: 1 cgggcaaacgctgaaa.....caactctggagctctctgag 271

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	112	41.3	733	12	BF674954
2	81.2	20.0	980	14	BQ551334
3	53	19.6	902	14	BQ550118
4	53	19.6	910	14	BQ548623
5	53	19.6	935	14	BQ650042
6	40.6	15.0	600	12	BG803927

7	37	13.7	535	12	BG566944
8	37	13.7	592	12	BG562803
9	37	13.7	630	12	BG571704
10	37	13.7	650	12	BG564662
11	37	13.7	730	12	BG621542
12	37	13.7	785	12	BG567864
13	37	13.7	788	12	BF677811
14	37	13.7	1101	17	CNS006BP
15	36.6	13.5	1083	14	BQ058613
16	36	13.3	590	14	BQ0807326
17	35.6	13.1	1101	17	CNS00DRA
18	35.2	13.0	1087	13	BM458295
19	35.2	13.0	1201	17	CNS015WQ
20	34.8	12.8	709	9	AL525279
21	34.6	12.8	990	17	CNS078MN
22	34.6	12.8	1026	13	BM044976
23	34.4	12.7	357	14	N75549
24	34	12.5	780	12	BF981904
25	33.8	12.5	368	9	AA565892
26	33.8	12.5	902	17	CNS00462
27	33.6	12.4	812	14	BQ229858
28	33.6	12.4	1047	14	BQ944238
29	33.6	12.4	1059	13	BM470242
30	33.6	12.4	1068	13	BM544574
31	33.4	12.3	507	12	BF754755
32	33.4	12.3	970	17	CNS06UR1
33	33.2	12.3	482	17	AZ854111
34	33.2	12.3	562	12	BF079717
35	33.2	12.3	845	9	AL571180
36	33.2	12.3	1096	14	BM913074
37	33	12.2	266	12	BG182293
38	33	12.2	272	12	BF888447
39	33	12.2	897	17	AZ185982
40	33	12.2	941	13	BM451198
41	33	12.2	970	9	AL545329
42	33	12.2	1087	17	CNS053SI
43	32.8	12.1	574	17	AQ370106
44	32.8	12.1	671	17	AQ395252
45	32.8	12.1	724	13	BI822441

ALIGNMENTS

RESULT 1  
BF674954  
LOCUS 602138032F1 NIH\_MGC\_83 Homo sapiens cDNA clone IMAGE:4274550 5',  
DEFINITION 733 bp mRNA linear EST 21-DEC-2000  
BF674954 mRNA sequence.

ACCESSION BF674954  
VERSION BF674954.1 GI:11948849  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 733)

NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)

Tissue Procurement: CLONETECH Laboratories, Inc.  
CDNA Library Preparation: CLONETECH Laboratories, Inc.

CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>  
Plate: L10K1091 row: p column: 07

High quality sequence stop: 649.  
Location/Qualifiers

FEATURES  
1. .733  
source



```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4274550"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCCGCGGCGGCACATG-3' (30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT      196 a      186 c      207 g      144 t
ORIGIN

Query Match      41.3%; Score 112; DB 12; Length 733;
Best Local Similarity 100.0%; Pred. No. 6.4e-23;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 CAGCCTTGCTCCTCACTACCCCACTCCAGAGGAGCAGCTATTATTAAGGGGAGCAG 219
Db 53 CAGCCTTGCTCCTCACTACCCCACTCCAGAGGAGCAGCTATTATTAAGGGGAGCAG 112

QY 220 GAGTGCAACAACAAAGACGGCCTGGGATACAACTCTGGAGTCTCTGAG 271
Db 113 GAGTGCAACAACAAAGACGGCCTGGGATACAACTCTGGAGTCTCTGAG 164

RESULT 2
BQ651334
LOCUS      BQ651334      980 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8500353 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6297803
5', mRNA sequence.
ACCESSION  BQ651334
VERSION    BQ651334.1 GI:21775506
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 980)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2507 row: n column: 12
            High quality sequence stop: 538.
            Location/Qualifiers
                1..980
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:6297803"
                /clone_lib="NIH_MGC_100"
                /tissue_type="hepatocellular carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGACAGG(G). Size-selected >500bp for average insert size
                1.8kb. Library constructed by Ling Hong in the laboratory
                of Gerald M. Rubin (University of California, Berkeley)
                using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                using ZAP-cDNA synthesis kit (Stratagene) and Superscript
```

```
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT      258 a      279 c      244 g      199 t
ORIGIN

Query Match      30.0%; Score 81.2; DB 14; Length 980;
Best Local Similarity 96.5%; Pred. No. 1.2e-13;
Matches 83; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 186 CTCCCGAGGAGGAGCAGCTATTTAAGGGGAGCAGAGTGAGCAACAAGACGGGCTG 245
Db 19 CCAACACAGGAGGAGCAGCTATTTAAGGGGAGCAGAGTGAGCAACAAGACGGGCTG 78

QY 246 GGATACAACTCTGGAGTCTCTGAG 271
Db 79 GGGATACAACTCTGGAGTCTCTGAG 104

RESULT 3
BQ650118
LOCUS      BQ650118      902 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8303325 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6270909
5', mRNA sequence.
ACCESSION  BQ650118
VERSION    BQ650118.1 GI:21774290
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 902)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2446 row: m column: 22
            High quality sequence stop: 692.
            Location/Qualifiers
                1..902
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:6270909"
                /clone_lib="NIH_MGC_100"
                /tissue_type="hepatocellular carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGACAGG(G). Size-selected >500bp for average insert size
                1.8kb. Library constructed by Ling Hong in the laboratory
                of Gerald M. Rubin (University of California, Berkeley)
                using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                II RT (Life Technologies). Note: this is a NIH_MGC
                Library."
BASE COUNT      236 a      266 c      211 g      189 t
ORIGIN

Query Match      19.6%; Score 53; DB 14; Length 902;
Best Local Similarity 100.0%; Pred. No. 3.3e-05;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 GGAGTGCAACAACAAAGACGGCCTGGGATACAACTCTGGAGTCTCTGAG 271
Db 1 GGAGTGCAACAACAAAGACGGCCTGGGATACAACTCTGGAGTCTCTGAG 53
```

```

RESULT 4
LOCUS      BQ648623                910 bp    mRNA          linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8297960 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6270023
           5', mRNA sequence.
ACCESSION  BQ648623
VERSION     BQ648623.1 GI:21772795
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 910)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LCM2444 row: h column: 24
            High quality sequence stop: 605.
            Location/Qualifiers
              1..910
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:6270023"
                /clone_lib="NIH MGC 100"
                /tissue_type="hepatocellular carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGACACGAG(G). Size-selected >500bp for average insert size
                1.8kb. Library constructed by Ling Hong in the laboratory
                of Gerald M. Rubin (University of California, Berkeley)
                using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                II RT (Life Technologies). Note: this is a NIH_MGC
                Library."
            BASE COUNT  237 a  265 c  224 g  184 t
            ORIGIN
            source
            Query March 19.6%; Score 53; DB 14; Length 910;
            Best Local Similarity 100.0%; Pred. No. 3.3e-05;
            Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XY  219 GGAGTGCAGAACAAACAGCGGCTGGGATACAACTCTGGAGTCTCTGAG 271
      1 GGAGTGCAGAACAAACAGCGGCTGGGATACAACTCTGGAGTCTCTGAG 53

      RESULT 5
      LOCUS      BQ650042                935 bp    mRNA          linear      EST 15-JUL-2002
      DEFINITION AGENCOURT_8302541 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271292
      5', mRNA sequence.
      ACCESSION  BQ650042
      VERSION     BQ650042.1 GI:21774214
      KEYWORDS   EST.
      SOURCE     human.
      ORGANISM   Homo sapiens
                  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
      REFERENCE  1 (bases 1 to 935)
      AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
      TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
      JOURNAL   Unpublished (1999)

      RESULT 6
      LOCUS      BQ803927                600 bp    mRNA          linear      EST 20-DEC-2001
      DEFINITION Q243-51 Mouse E14.5 retina lambda ZAP II Library Mus musculus CDNA,
      mRNA sequence.
      ACCESSION  BQ803927
      VERSION     BQ803927.1 GI:17950840
      KEYWORDS   EST.
      SOURCE     house mouse.
      ORGANISM   Mus musculus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      REFERENCE  1 (bases 1 to 600)
      AUTHORS   Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
                  White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
      TITLE     Gene expression in the developing mouse retina by EST sequencing
                  and microarray analysis
      JOURNAL   Nucleic Acids Res. 29 (24), 4993-4993 (2001)
      MEDLINE    21671825
      COMMENT    Contact: Klein WH
                  Department of Biochemistry and Molecular Biology
                  University of Texas M.D. Anderson Cancer Center
                  Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
                  Tel: 713 732 3646
                  Fax: 713 790 0329,
                  Location/Qualifiers
                    1..600
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
                      /tissue_type="neural retina"
                      /dev_stage="embryonic day 14.5 post-fertilization"
                      /note="Vector: pAMPl0 (Gibco); Cloned unidirectionally.

```

```

COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LCM2447 row: m column: 21
            High quality sequence stop: 691.
            Location/Qualifiers
              1..935
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:6271292"
                /clone_lib="NIH_MGC_100"
                /tissue_type="hepatocellular carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGACACGAG(G). Size-selected >500bp for average insert size
                1.8kb. Library constructed by Ling Hong in the laboratory
                of Gerald M. Rubin (University of California, Berkeley)
                using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                II RT (Life Technologies). Note: this is a NIH_MGC
                Library."
            BASE COUNT  261 a  257 c  226 g  191 t
            ORIGIN
            Query March 19.6%; Score 53; DB 14; Length 935;
            Best Local Similarity 100.0%; Pred. No. 3.3e-05;
            Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  219 GGAGTGCAGAACAAACAGCGGCTGGGATACAACTCTGGAGTCTCTGAG 271
      1 GGAGTGCAGAACAAACAGCGGCTGGGATACAACTCTGGAGTCTCTGAG 53

      Db 1 GGAGTGCAGAACAAACAGCGGCTGGGATACAACTCTGGAGTCTCTGAG 53

      RESULT 6
      LOCUS      BQ803927                600 bp    mRNA          linear      EST 20-DEC-2001
      DEFINITION Q243-51 Mouse E14.5 retina lambda ZAP II Library Mus musculus CDNA,
      mRNA sequence.
      ACCESSION  BQ803927
      VERSION     BQ803927.1 GI:17950840
      KEYWORDS   EST.
      SOURCE     house mouse.
      ORGANISM   Mus musculus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      REFERENCE  1 (bases 1 to 600)
      AUTHORS   Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
                  White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
      TITLE     Gene expression in the developing mouse retina by EST sequencing
                  and microarray analysis
      JOURNAL   Nucleic Acids Res. 29 (24), 4993-4993 (2001)
      MEDLINE    21671825
      COMMENT    Contact: Klein WH
                  Department of Biochemistry and Molecular Biology
                  University of Texas M.D. Anderson Cancer Center
                  Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
                  Tel: 713 732 3646
                  Fax: 713 790 0329,
                  Location/Qualifiers
                    1..600
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
                      /tissue_type="neural retina"
                      /dev_stage="embryonic day 14.5 post-fertilization"
                      /note="Vector: pAMPl0 (Gibco); Cloned unidirectionally.

```

Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps (Mannatis); Cloning Technique: CUA Cloning (CloneAmp, Life Technologies); Average insert size: 1.8 Kb; Insertion site: TAGCTCCATGAATCTGAGT---. Other information regarding entire library may be found at [http://psa.swmed.edu/Data/Libraries/microarray\\_cdna\\_library.htm](http://psa.swmed.edu/Data/Libraries/microarray_cdna_library.htm).

BASE COUNT	165 a	167 c	181 g	87 t
ORIGIN				

Query Match	15.0%;	Score 40.6;	DB 12;	Length 600;
Best Local Similarity	56.3%;	Pred. No. 0.15;		
Matches	76;	Conservative 0;	Mismatches 59;	Indels 0;
QY	63	CTCAACTGTGCTGCCAGCTGATGAGGGCAAGGAAGGATTACCTAGGGGTATGGGCG	122	
Db	101	CTCAGCTGTGCTCTGGAGGAAGCGACTGGAGCAGAGAGTTCGTGGGGTTGGCAGTG	160	
QY	123	ACCAATCTCTGTGATCCACCAACTGACACGCGCCCATCCCCAGCCTTGTGCCTCACCTTACCCC	182	
Db	161	ACCGACACTGAGGTCCTTCTGGCCCCCCCCCCCCCCCCCGCAGAGCCTCCTTCATGTGACCCA	220	
QY	183	CAACCTCCGAGAGG	197	
Db	221	GAGACGGCAAGGG	235	

```

RESULT 7
BG566944 linear EST 10-APR-2001
LOCUS 60258944BP1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723810 5',
DEFINITION mRNA sequence.
ACCESSION BG566944
VERSION BG566944.1 GI:13574597
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 535)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
JOURNAL
COMMENT

```

```

FEATURES
source
high quality sequence setup: 553.
Location/Qualifiers
1. 535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4723810"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pBMR-LIB (Clontech); Site 1: Sfil (ggcgctctcgcc); Site 2: Sfil (ggcattatgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGCGCATTAATGCC-3', adaptor sequence: 5'-ATTCTAGAGCGGAGGGCGGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
146 a 130 c 150 g 109 t
BASE COUNT
ORIGIN

```

```

Query Match      13.7%   Score 37;   DB 12;   Length 535;
Best Local Similarity 100.0%;   Pred. No. 1.8;
Matches 37;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```

RESULT 8	BG562803	linear	EST 10-APR-2001
LOCUS	602581630F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4709527 5',	mRNA	
DEFINITION	MRCNA sequence.		
ACCESSION	BG562803		
VERSION	BG562803.1 GI:13570455		
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (Bases 1 to 592)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: csapbs@email.nih.gov Tissue Procurement: CLONETECH Laboratories, Inc. cDNA Library Preparation: CLONETECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through: the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLCMI546 row: 1 column: 08 High quality sequence stop: 592.		

```

BASE COUNT      159 a    147 c    163 g    123 t
ORIGIN
Query Match          13.7% ; Score 37; DB 12; Length 592;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY   235  AAGACGGCCTGGGATACAACACTCGGAGTCTCTTGAG 271
        |||||
Db    2    AAGACGGCCTGGGATACAACACTCGGAGTCTCTTGAG 38
        |||||

RESULT 9
BG573704
LOCUS
DEFINITION
mRNA sequence.
BG573704
602594/753P1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:472133 5' linear EST 10-APR-2001
BG573704 630 bp mRNA
BG573704.1 GI:13581357

```

RESULT 9	EG573704	LOCUS	DEFINITION	EG573704	630 bp	mRNA	linear	EST 10-APR-2001
				CG2594753F1	NIH_MGC_79	Homo sapiens	cDNA clone	IMAGE:472133.5,
				602594753F1				
				RNA sequence.				
		ACCESSION		EG573704				
		VERSION		EG573704.1				
				GI:13581357				



insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

BASE COUNT 196 a 195 c 184 g 155 t  
ORIGIN

Query Match 13.7%; Score 37; DB 12; Length 730;

Best Local Similarity 100.0%; Pred. No. 1.9; Mismatches 0; Indels 0; Gaps 0;

QY 235 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 271

DB 3 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 39

# RESULT 12

BG567864

LOCUS BG567864 785 bp mRNA linear EST 10-APR-2001  
DEFINITION 60258668F1 NIH\_MGC\_76 Homo sapiens cDNA clone IMAGE:4715612 5', mRNA sequence.

ACCESSION BG567864

VERSION BG567864.1 GI:13575517

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 785)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI562 row: i column: 21

High quality sequence stop: 764.

Location/Qualifiers

FEATURES

source

1. .785

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4715612"

/clone\_lib="NIH\_MGC\_76"

/lab\_host="DH10B (TI phage-resistant)"

/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:

SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); 5' and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CAGGCGCATATGGC-3' and 3' adaptor sequence:

5'-ATTCTAGGCGGCGGCGGCACATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.85

kb (range 1.0-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH\_MGC Library."

BASE COUNT 205 a 225 c 187 g 168 t

# ORIGIN

Query Match 13.7%; Score 37; DB 12; Length 785;

Best Local Similarity 100.0%; Pred. No. 2; Mismatches 0; Indels 0; Gaps 0;

QY 235 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 271

DB 3 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 39

# RESULT 13

BF677811

LOCUS

DEFINITION

60208532F1 NIH\_MGC\_83 Homo sapiens cDNA clone IMAGE:4249432 5', mRNA sequence.

ACCESSION BF677811

VERSION BF677811.1 GI:11951706

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 788)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI069 row: 1 column: 05

High quality sequence stop: 668.

Location/Qualifiers

FEATURES

source

1. .788

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4249432"

/clone\_lib="NIH\_MGC\_83"

/lab\_host="DH10B (TI phage-resistant)"

/note="Organ: prostate; Vector: pDNR-LIB (Clontech);

Site 1: SfiI (ggccctcgcc); Site 2: SfiI (ggccattatggcc)

; 5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CAGGCGCATATGGC-3' and 3' adaptor

sequence: 5'-ATTCTAGGCGGCGGCGGCACATG-dt(30)BN-3'

(where B = A, C, G, or T). Average insert size 1.4 kb

(range 0.5-4.0 kb). 14/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA)."

BASE COUNT 207 a 203 c 203 g 175 t

# ORIGIN

Query Match 13.7%; Score 37; DB 12; Length 788;

Best Local Similarity 100.0%; Pred. No. 2; Mismatches 0; Indels 0; Gaps 0;

QY 235 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 271

DB 2 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 38

# RESULT 14

CNS006BP

LOCUS

DEFINITION

CNS006BP 1101 bp DNA linear GSS 03-JUN-1999

Drosophila melanogaster genome survey sequence TET3 end of BAC #

BACR13018 of RPI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL064052

AL064052.1 GI:4944232

KEYWORDS GSS.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

TITLE

```

1. 1083
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5814496"
/clone_lib="NIH MGC 99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into BcoRI/XhoI sites using the following 5' adaptor:
GGACACGAG(G). Size-selected >500bp for average insert size
1 kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
253 a 325 c 319 g 184 t 2 others
BASE COUNT
ORIGIN

```

```

Query Match      13.5%  Score 36.6;  DB 14;  Length 1083;
Best Local Similarity 53.6%  Pred. No. 2.9;
Matches 75;  Conservative 0;  Mismatches 65;  Indels 0;  Gaps 0;

QY      56  CCCAATCCTCAACTGTGCTCTGCGCAGCTGATGAGGGGAGGAAAGGGATTACCTAGGGGT 115
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      871  CCCCAGTCCCTACCCCTTCTCTCCCATGAAAGGGGTGGGCCAGGGGCGNCCTGGGGT 930

QY      116  ATGGGCGACCAATCTGAGTCCCAACTGACCACGCCCATCCCCAGCCTTGTGCTCAC 175
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      931  ATCCATGTGTCACCTTGCGCCCAAGACAGAGAACCCCTTGGGCTCCCGCTGGCCCTCCC 990

QY      176  CTACCCCAACCTCCAGAG 195
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      991  CTTGCCCTTTTTCCTGGG 1010

Search completed: February 20, 2003, 01:09:01
Job time : 1325.61 secs

```

Search completed: February 20, 2003, 01:09:01  
Job time : 1325.61 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

M nucleic - nucleic search, using sw model  
un on: February 19, 2003, 19:28:33 ; Search time 1032.56 Seconds  
(without alignments)  
9357.426 Million cell updates/sec  
title: US-09-808-388-6  
effect score: 332  
sequence: 1 gacacattcgacaaaacta.....caactctggagctctctgag 332

coring table: IDENTITY NUC  
Gapop 10.0 , Capext 1.0  
eached: 2054640 seqs, 14551402878 residues  
otal number of hits satisfying chosen parameters: 4109280

inimum DB seq length: 0  
aximum DB seq length: 2000000000  
ost-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.man.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	332	100.0	332	6	AX251578	AX251578 Sequence
2	271	81.6	271	6	AX251577	AX251577 Sequence
c 3	268.4	80.8	194247	2	AL358253	AL358253 Homo sapi
c 4	268.2	80.8	11116	6	AX015387	AX015387 Sequence
c 5	268.2	80.8	1268	6	AX015532	AX015532 Sequence
c 6	213	64.2	1080	6	I09231	I09231 Sequence 36
c 7	213	64.2	1080	9	HUMRASF1	M2429 Human RASF-
c 8	171	51.5	157470	2	AL360079	AL360079 Homo sapi
c 9	165	49.7	5356	6	AX348874	AX348874 Sequence
c 10	165	49.7	5728	6	AX344469	AX344469 Sequence
c 11	165	49.7	6083	6	AX251472	AX251472 Sequence
c 12	150.2	45.2	5356	6	AX348873	AX348873 Sequence
c 13	150.2	45.2	5728	6	AX344468	AX344468 Sequence
c 14	150.2	45.2	6083	6	AX251471	AX251471 Sequence
c 15	118.8	35.8	997	9	BC005919	BC005919 Homo sapi
c 16	84.4	25.4	3330	6	AX401819	AX401819 Sequence
c 17	84.4	25.4	3330	10	RNPPLA2G	AF375595 Rattus no
c 18	83.2	25.1	2697	10	AF375595	M37127 Rat group I
c 19	82.8	24.9	3366	10	RATGIIPHOS	X52613 Rat DNA for
c 20	60.6	18.3	497	10	RNPPLA2	AC118094 Rattus no
c 21	56.6	17.0	157317	2	AC118094	U32313 Mus musculu
c 22	49.6	14.9	4438	10	MMU32313	AC002108 Genomic S
c 23	49.6	14.9	41125	10	AC002108	AL844178 Mus muscu
c 24	49.6	14.9	186589	2	AL844178	AX251575 Sequence
c 25	41	12.3	41	6	AX251575	AL357060 Human DNA
c 26	39.8	12.0	182209	9	AL357060	AC111113 Mus muscu
c 27	38.2	11.5	183647	2	AC111113	AF429315 Homo sapi
c 28	37.8	11.4	125020	9	AF429315	G67426 D7S3095 RPC
c 29	37.6	11.3	339	11	G67426	AP000373 Homo sapi
c 30	37.2	11.2	100000	9	AP000073	AC103170 Rattus no
c 31	36.2	10.9	170466	2	AC103170	AC129602 Mus muscu
c 32	35.6	10.7	201441	2	AC129602	AC093984 Rattus no
c 33	35.6	10.7	206407	1	AC093984	AE013943 Versinia
c 34	35.2	10.6	13706	2	AE013943	AC108964 Rattus no
c 35	35.2	10.6	141706	2	AC108964	AJ414144 Versinia
c 36	35.2	10.6	208050	1	AJ414144	AL353609 Homo sapi
c 37	35	10.5	165867	2	AL353609	AL365274 Human DNA
c 38	35	10.5	171941	9	AL365274	AC107269 Rattus no
c 39	34.8	10.5	102051	2	AC107269	Continuation (6 of
c 40	34.6	10.4	34015	2	AL672265	AC128259 Rattus no
c 41	34.6	10.4	138647	2	AC128259	AC110977 Rattus no
c 42	34.6	10.4	173424	2	AC110977	AC119246 Mus muscu
c 43	34.4	10.4	46328	2	AC119246	AF429315 Homo sapi
c 44	34.4	10.4	125020	9	AF429315	AC119064 Sus scrofa
c 45	34.4	10.4	150366	2	AC119064	

ALIGNMENTS

RESULT 1  
AX251578  
LOCUS AX251578 332 bp DNA linear PAT 05-OCT-2001  
DEFINITION Sequence 6 from Patent WO0168845.  
ACCESSION AX251578  
VERSION AX251578.1 GI:15985001  
KEYWORDS synthetic construct,  
synthetic construct,  
artificial sequences.  
ORGANISM 1 (bases 1 to 332)  
REFERENCE Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Bereziat,G.  
AUTHORS Inflammation-inducible hybrid promoters, vectors containing same  
TITLE and uses thereof  
JOURNAL Patent: WO 0168845-A 6 20-SEP-2001;

```

FEATURES             source
    source            Location/Qualifiers
    1..332
    /organism="synthetic construct"
    /db_xref="taxon:32630"
    /note="promoteur hybride PPRE/PLA2s"
BASE COUNT          96 a 91 c 82 g 63 t
ORIGIN
Query Match          100.0%; Score 332; DB 6; Length 332;
Best Local Similarity 100.0%; Pred. No. 3.9e-97;
Matches 332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTACCAATTCGACAAACTAGGTCAAAGTCTCAAAACTAGGTCAAAGGTCAAATTGGA 60
DB 1 GTACCAATTCGACAAACTAGGTCAAAGTCTCAAAACTAGGTCAAAGGTCAAATTGGA 60
QY 61 ACGGGCAAACTGCTGAATGTGTTTGGATCAGCTACTGACACGTAGGTTTCCCA 120
DB 61 ACGGGCAAACTGCTGAATGTGTTTGGATCAGCTACTGACACGTAGGTTTCCCA 120
QY 121 ATCTCAACTCTGCTGCGAGCTGATGAGGGGAAGAAAGGATTACTAGGGGTATGG 180
DB 121 ATCTCAACTCTGCTGCGAGCTGATGAGGGGAAGAAAGGATTACTAGGGGTATGG 180
QY 181 GCGACCAATCTGAGTCCACCAACTGACACGCCCATCCCGAGCCTTGCTTACCTAC 240
DB 181 GCGACCAATCTGAGTCCACCAACTGACACGCCCATCCCGAGCCTTGCTTACCTAC 240
QY 241 CCCCACTCCAGAGGACAGCTATTAAGGGGAGGAGGTGACAGCAAAACAGAC 300
DB 241 CCCCACTCCAGAGGACAGCTATTAAGGGGAGGAGGTGACAGCAAAACAGAC 300
QY 301 GGCCTGGGGATACAACTCTGGAGTCTCTTGAG 332
DB 301 GGCCTGGGGATACAACTCTGGAGTCTCTTGAG 332

RESULT 2
AX251577             271 bp      DNA      linear      PAT 05-OCT-2001
LOCUS                AX251577
DEFINITION           Sequence 5 from Patent WO0168845.
ACCESSION            AX251577
VERSION              AX251577.1 GI:15985000
KEYWORDS              synthetic construct.
                     synthetic construct
                     artificial sequences.
ORIGIN
REFERENCE            Massaad,C., Berenbaum,F., Olivier,J.L., Salvat,C. and Berziat,G.
AUTHORS              Inflammation-inducible hybrid promoters, vectors containing same
TITLE                and uses thereof
JOURNAL              Patent: WO 0168845-A 5 20-SEP-2001;
                     Aventis Pharma S.A. (FR)
FEATURES             source
    source            Location/Qualifiers
    1..271
    /organism="synthetic construct"
    /db_xref="taxon:32630"
    /note="fragment du promoteur PLA2s"
BASE COUNT          70 a 79 c 71 g 51 t
ORIGIN
Query Match          91.6%; Score 271; DB 6; Length 271;
Best Local Similarity 100.0%; Pred. No. 3.3e-77;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 CGCGCAAACTGCTGAATGTGTTTGGATCAGCTACTGACAGTAAAGTTTCCCAA 121
DB 1 CGCGCAAACTGCTGAATGTGTTTGGATCAGCTACTGACAGTAAAGTTTCCCAA 60
QY 122 TCCTCAACTCTGCTGCGACCTGATGAGGGGAAGAAAGGATTACTAGGGGTATGG 181
DB 61 TCCTCAACTCTGCTGCGACCTGATGAGGGGAAGAAAGGATTACTAGGGGTATGG 120

QY 182 CGACCAATCTGAGTCCACCAACTGACCAAGCCATCCCGAGCCTTGCTTACCTACC 241
DB 121 CGACCAATCTGAGTCCACCAACTGACCAAGCCATCCCGAGCCTTGCTTACCTACC 180
QY 242 CCACACCTCCAGAGGAGAGCTATTATTAAGGGGAGGAGGTGACAGCAAAACAGAC 301
DB 181 CCACACCTCCAGAGGAGAGCTATTATTAAGGGGAGGAGGTGACAGCAAAACAGAC 240
QY 302 GCTGGGGATACAACTCTGGAGTCTCTTGAG 332
DB 241 GCTGGGGATACAACTCTGGAGTCTCTTGAG 271

AL358253             194247 bp      DNA      linear      HTG 25-JUL-2002
LOCUS                AL358253/c
DEFINITION           Homo sapiens chromosome 1 clone RP11-460G22, *** SEQUENCING IN
                     PROGRESS ***, in ordered pieces.
ACCESSION            AL358253
VERSION              AL358253.14 GI:21998163
KEYWORDS              HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE               Homo sapiens.
ORGANISM              Homo sapiens
                     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE            Bagdaley,C.
AUTHORS              Direct Submission
TITLE                Submitted (24-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
                     Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                     humbrey@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
                     On Jul 26, 2002 this sequence version replaced gi:21748207.
COMMENT              ----- Genome Center
                     Center: Wellcome Trust Sanger Institute
                     Center code: SC
                     Web site: http://www.sanger.ac.uk
                     Contact: humbrey@sanger.ac.uk
                     ----- Project Information
                     Center project name: BA460G22
                     ----- Summary Statistics
                     Assembly program: XGAP4; version 4.5
                     Chemistry: Dye-terminator; 0% of reads
                     Chemistry: Dye-terminator-amersham; 1% of reads
                     Chemistry: Dye-terminator ET-amersham; 3% of reads Chemistry:
                     Dye-terminator Big Dye; 95% of reads
                     Consensus quality: 194218 bases at least Q40
                     Consensus quality: 194247 bases at least Q30
                     Consensus quality: 194247 bases at least Q20
                     Insert size: 194247; sum-of-contigs
                     Insert size: 187546; 12.3% error; agarose-fp
                     Quality coverage: 10.76x in Q20 bases; sum-of-contigs Quality
                     coverage: 12.13x in Q20 bases; agarose-fp
                     -----
                     * NOTE: This is a 'working draft' sequence.
                     * This sequence will be replaced
                     * by the finished sequence as soon as it is available and
                     * the accession number will be preserved.
FEATURES             Location/Qualifiers
    source            1..194247
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-460G22"
                     /clone_lib="RPCI-11.2"
                     1..194247
                     /note="assembly fragment:05625"
BASE COUNT          52994 a 42433 c 43263 g 55557 t
ORIGIN
Query Match          80.8%; Score 268.4; DB 2; Length 194247;
Best Local Similarity 97.8%; Pred. No. 2.8e-76;
Matches 272; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```



RESULT 5  
LOCUS AX015532/c 1268 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 226 from Patent WO9951727.  
ACCESSION AX015532  
VERSION AX015532.1 GI:10041414  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1268)  
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and  
Pilarsky,C.  
TITLE Human nucleic acid sequences of normal ovary tissue  
JOURNAL Patent: WO 9951727-A 226 14-OCT-1999;  
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN  
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN  
(DE); PILARSKY CHRISTIAN (DE)  
FEATURES  
source Location/Qualifiers  
1..1268  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 262 a 313 c 343 g 350 t  
ORIGIN  
Query Match 80.8%; Score 268.2; DB 6; Length 1268;  
Best Local Similarity 98.9%; Pred. No. 2.8e-76;  
Matches 270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 60 AACGGCGGCAAACTGCTGAAATGTTGTTGGCATCAGCTACTGACACGTAAGGTTTCCC 119  
Db 1267 AAGACTGCAAAATGCTCTGAATGTTGTTGGCATCAGCTACTGACACGTAAGGTTTCCC 1208  
QY 120 AATCTCAACTCTGCTGCCAGCTGATAGGGGAGGAAAGGATTAACCTAGGGGTATG 179  
Db 1207 AATCTCAACTCTGCTGCCAGCTGATAGGGGAGGAAAGGATTAACCTAGGGGTATG 1148  
QY 180 GGCGACCAATCTGAGTCCACCACTGACACGCCCATCCCCAGCCTTGTCCTCACCTA 239  
Db 1147 GGCGACCAATCTGAGTCCACCACTGACACGCCCATCCCCAGCCTTGTCCTCACCTA 1088  
QY 240 CCCCCAACCTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGAGTGCGAGAACAAAGA 299  
Db 1087 CCCCCAACCTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGAGTGCGAGAACAAAGA 1028  
QY 300 CGGCTGGGGATACAACTCTGGAGTCTCTGAG 332  
Db 1027 CGGCTGGGGATACAACTCTGGAGTCTCTGAG 995  
RESULT 6  
LOCUS I09231 1080 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 36 from Patent WO 8901773.  
ACCESSION I09231  
VERSION I09231.1 GI:588062  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1080)  
AUTHORS Johnson,L.K., Seilhamer,J.J., Pruzanski,W. and Vadas,P.  
TITLE SYNOPSIS PHOSPHOLIPASES  
JOURNAL Patent: WO 8901773-A 36 09-MAR-1989;  
FEATURES  
source Location/Qualifiers  
1..1080  
/organism="unknown"  
BASE COUNT 306 a 242 c 303 g 229 t  
ORIGIN  
Query Match 64.2%; Score 213; DB 6; Length 1080;  
Best Local Similarity 93.1%; Pred. No. 3.1e-58;

Y 55 ATTGCAACGGCGGCAAACTGCTGAAATGTTGTTGGCATCAGCTACTGACACGTAAGGT 114  
b 165734 ATGGAAGACTGCAAAATGCTGAAATGTTGTTGGCATCAGCTACTGACACGTAAGGT 165675  
Y 115 TTCCCAATCTCTCAACTCTGCTGCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGGG 174  
b 165674 TTCCCAATCTCTCAACTCTGCTGCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGGG 165615  
Y 175 GTATGGGCGCAATCTGAGTCCACCACTGACACGCCCATCCCGCCTTGTCCTC 234  
b 165614 GTATGGGCGCAATCTGAGTCCACCACTGACACGCCCATCCCGCCTTGTCCTC 165555  
Y 235 ACCTACCCCACTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGATGCGACAA 294  
b 165554 ACCTACCCCACTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGATGCGACAA 165495  
Y 295 CAAGACGCCCTGGGGATACAACTCTGAGTCTCTGAG 332  
b 165494 CAAGACGCCCTGGGGATACAACTCTGAGTCTCTGAG 165457  
RESULT 4  
LOCUS X015387 1116 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 53 from Patent WO9951727.  
ACCESSION X015387  
VERSION X015387.1 GI:10041367  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1116)  
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and  
Pilarsky,C.  
TITLE Human nucleic acid sequences of normal ovary tissue  
JOURNAL Patent: WO 9951727-A 53 14-OCT-1999;  
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN  
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN  
(DE); PILARSKY CHRISTIAN (DE)  
FEATURES  
source Location/Qualifiers  
1..1116  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 311 a 311 c 266 g 228 t  
ORIGIN  
Query Match 80.8%; Score 268.2; DB 6; Length 1116;  
Best Local Similarity 98.9%; Pred. No. 2.8e-76;  
Matches 270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 60 AACGGCGGCAAACTGCTGAAATGTTGTTGGCATCAGCTACTGACACGTAAGGTTTCCC 119  
Db 2 AAGACTGCAAAATGCTGAAATGTTGTTGGCATCAGCTACTGACACGTAAGGTTTCCC 61  
QY 120 AATCTCAACTCTGCTGCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGGGGTATG 179  
Db 62 AATCTCAACTCTGCTGCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGGGGTATG 121  
QY 180 GGCGACCAATCTGAGTCCACCACTGACACGCCCATCCCCAGCCTTGTCCTCACCTA 239  
Db 122 GGCGACCAATCTGAGTCCACCACTGACACGCCCATCCCCAGCCTTGTCCTCACCTA 181  
QY 240 CCCCCAACCTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGATGCGAGAACAAAGA 299  
Db 182 CCCCCAACCTCCAGAGGAGCAGCTATTAAAGGGGAGCAGGATGCGAGAACAAAGA 241  
QY 300 CGGCTGGGGATACAACTCTGAGTCTCTGAG 332  
Db 242 CGGCTGGGGATACAACTCTGAGTCTCTGAG 274

Matches 256; Conservative 0; Mismatches 15; Indels 4; Gaps 3;

QY 61 ACGGGCAAACTGCTGAAATGTGTTTGTCATCAGCTACTGACAGTAAGG-TTTCCC 119

Db 762 ACTGGCAAACTGCTGAAATGTGTTTGGCATCAGGCTACTGACAGTAAGGTTTCCC 821

QY 120 AATCTCTCAACTCTGTCCTG--CCAGCTGATGAGGGGAAGGAAAGGATTACCTAGGGGTA 177

Db 822 AATCTCTCAACTCTGTCCTGGCAGGCTGATGAGGGGAAGGAAAGGATTACCTAGGGGTA 881

QY 178 TGGGGCAAACTCTGAGTGCACCACTGACAGCCATCCAGGCTTGTGCTTCACC 237

Db 882 TGGGGCAAACTCTGAGTGCACCACTGACAGCCATCCAGGCTTGTGCTTCACC 941

QY 238 TACCCCACTCTCCAGAGGGAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAA 297

Db 942 TACCCCACTCTCCAGAGGGAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAA 1000

QY 298 GACGGCTGGGATACACTCTGGAGTCTCTCTGAG 332

Db 1001 GACGGCTGGGATACACTCTGGAGTCTCTCTGAG 1035

RESULT 7

HUMRASFAL

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

AUTHORS

TITLE

COMMENT

FEATURES

source

intron

BASE COUNT

ORIGIN

Query Match 64.2%; Score 213; DB 9; Length 1080;

Best Local Similarity 93.1%; Pred. No. 3.1e-58;

Matches 256; Conservative 0; Mismatches 15; Indels 4; Gaps 3;

QY 61 ACGGGCAAACTGCTGAAATGTGTTTGTCATCAGCTACTGACAGTAAGG-TTTCCC 119

Db 762 ACTGGCAAACTGCTGAAATGTGTTTGGCATCAGGCTACTGACAGTAAGGTTTCCC 821

QY 120 AATCTCTCAACTCTGTCCTG--CCAGCTGATGAGGGGAAGGAAAGGATTACCTAGGGGTA 177

Db 822 AATCTCTCAACTCTGTCCTGGCAGGCTGATGAGGGGAAGGAAAGGATTACCTAGGGGTA 881

QY 178 TGGGGCAAACTCTGAGTGCACCACTGACAGCCATCCAGGCTTGTGCTTCACC 237

Db 882 TGGGGCAAACTCTGAGTGCACCACTGACAGCCATCCAGGCTTGTGCTTCACC 941

QY 238 TACCCCACTCTCCAGAGGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGAACAAACAA 297

Db 942 TACCCCACTCTCCAGAGGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGAACAAACAA 1000

QY 298 GACGGCTGGGATACACTCTGGAGTCTCTCTGAG 332

Db 1001 GACGGCTGGGATACACTCTGGAGTCTCTCTGAG 1035

RESULT 8

AL360079/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL360079 157470 bp DNA linear HTG 10-JUL-2001

Homo sapiens chromosome 1 clone RP11-66M4, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 22 unordered pieces.

AL360079

AL360079.3 GI:9801103

HTG; HTGS\_PHASE1; HTGS\_CANCELLED.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 157470)

McLay, K.

Direct Submission

Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonesrequest@sanger.ac.uk

On Aug 14, 2000 this sequence version replaced gi:8919533.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA68M4

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator ET-amersham; 1% of reads

Dye-terminator Big Dye, 98% of reads

Consensus quality: 145607 bases at least Q40

Consensus quality: 150854 bases at least Q30

Consensus quality: 153362 bases at least Q20

Insert size: 155370; sum-of-contigs

Insert size: 175968; 2.3% error; agarose-fp

Quality coverage: 3.40x in Q20 bases; sum-of-contigs Quality coverage: 3.16x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 15108: contig of 15108 bp in length

15109 15208: gap of 100 bp

15209 20077: contig of 4869 bp in length

20078 20177: gap of 100 bp

20178 26598: contig of 6421 bp in length

26599 26698: gap of 100 bp

26699 31874: contig of 5176 bp in length

31875 31974: gap of 100 bp

31975 39580: contig of 7606 bp in length

39581 39680: gap of 100 bp

39681 62977: contig of 23297 bp in length

62978 63077: gap of 100 bp

63078 75654: contig of 12577 bp in length

75655 75754: gap of 100 bp

75755 90667: contig of 14913 bp in length

90668 90767: gap of 100 bp

90768 93945: contig of 3178 bp in length

```
* 93946 94045: gap of 100 bp
* 94046 95022: contig of 4977 bp in length
* 99023 99122: gap of 100 bp
* 99123 101699: contig of 2577 bp in length
* 101700 101799: gap of 100 bp
* 101800 105241: contig of 3442 bp in length
* 105242 105341: gap of 100 bp
* 105342 107538: contig of 2197 bp in length
* 107539 107638: gap of 100 bp
* 107639 111008: contig of 3370 bp in length
* 111009 111108: gap of 100 bp
* 111109 113755: contig of 2647 bp in length
* 113756 113855: gap of 100 bp
* 113856 121296: contig of 7441 bp in length
* 121297 121386: gap of 100 bp
* 121387 125841: contig of 4445 bp in length
* 125842 125941: gap of 100 bp
* 125942 129256: contig of 3315 bp in length
* 129257 129356: gap of 100 bp
* 129357 136269: contig of 6913 bp in length
* 136270 136369: gap of 100 bp
* 136370 138879: contig of 2510 bp in length
* 138880 138979: gap of 100 bp
* 138980 144504: contig of 5525 bp in length
* 144505 144604: gap of 100 bp
* 144605 157470: contig of 12866 bp in length.

FEATURES
    Location/Qualifiers
        1..157470
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /clone="RP11-66M4"
            /clone_lib="RPC1-11.1"
            1..15108
                /note="assembly_fragment:00413"
                clone_end.SP6
                vector_side:left
            15209..20077
                /note="assembly_fragment:00153"
            20178..26598
                /note="assembly_fragment:00168"
            26699..31874
                /note="assembly_fragment:00204"
            31975..339580
                /note="assembly_fragment:00207"
            39681..62377
                /note="assembly_fragment:00211"
            63078..75654
                /note="assembly_fragment:00223"
            75755..90667
                /note="assembly_fragment:00330"
            90768..93945
                /note="assembly_fragment:00333"
            94046..99022
                /note="assembly_fragment:00365"
            99123..101699
                /note="assembly_fragment:00421"
            101800..105241
                /note="assembly_fragment:00566"
            105342..107538
                /note="assembly_fragment:00589"
            107639..111008
                /note="assembly_fragment:00674"
            111109..113755
                /note="assembly_fragment:00850"
            113856..121296
                /note="assembly_fragment:01043"
            121397..125841
                /note="assembly_fragment:01136"
            125942..129256
                /note="assembly_fragment:01206"
            129357..136269
                /note="assembly_fragment:01276"
```

```
misc_feature 136370..138879
    /note="assembly_fragment:01282"
misc_feature 138980..144504
    /note="assembly_fragment:01387"
misc_feature 144605..157470
    /note="assembly_fragment:01404"
BASE COUNT 43730 a 34140 c 34257 g 43227 t 2116 others
ORIGIN

Query Match 51.5%; Score 171; DB 2; Length 157470;
Best Local Similarity 94.7%; Pred. No. 1.8e-44; Indels 0; Gaps 0;
Matches 177; Conservative 0; Mismatches 10;

QY 55 ATTGGAACGGCGCAAACTGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGT 114
    |||||
Db 15395 ATGGAACACTGCAAACTGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGT 15336

QY 115 TTCCCAATCTCAACTCTGCTGCGCAGCTGATGAGGGNAGGAAAGGNTTACCTAGG 174
    |||||
Db 15335 TTCCCAATCTCAACTCTGCTGCGCAGCTAATGAGGGGAAGAAAGGATACCTAGG 15276

QY 175 GTATGGGGACCAATCTCTGAGTCCACCACTGACCAAGCCCATCCCGAGCTTGTGCCTC 234
    |||||
Db 15275 GTGTTGGGACCATCTCTGAGTCCACCACTGACCAAGCCCATCCCGAGCTTGTGCCTC 15216

QY 235 ACCTACC 241
    |||||
Db 15215 ACCTACC 15209

RESULT 9
AX348874/c
LOCUS AX348874 5356 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 332 from Patent WO0202807.
ACCESSION AX348874
VERSION AX348874.1 GI:18614909
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM synthetic construct.
            artificial sequences.
REFERENCE 1
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with cell signalling
JOURNAL Patent: WO 0202807-A 332 10-JAN-2002;
Epigenomics AG (DE)
FEATURES
    Location/Qualifiers
        1..5356
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 1357 a 60 c 1167 g 2772 t
ORIGIN

Query Match 49.7%; Score 165; DB 6; Length 5356;
Best Local Similarity 74.7%; Pred. No. 1.5e-42; Indels 0; Gaps 0;
Matches 207; Conservative 0; Mismatches 70;

QY 55 ATTGGAACGGCGCAAACTGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGT 114
    |||||
Db 1336 ATAAAACACTACAACACTACTAAATATATTTTACATCAACTACTACTACGTAAGT 1277

QY 115 TTCCCAATCTCAACTCTGCTGCGCAGCTGATGAGGGNAGGAAAGGATACCTAGG 174
    |||||
Db 1276 TTCCCAATCTCAACTCTGCTGCGCAGCTAATATAAAAAAATAATACCTAAAA 1217

QY 175 GTATGGGGACCAATCTCTGAGTCCACCACTGACCAAGCCCATCCCGAGCTTGTGCCTC 234
    |||||
Db 1216 ATATAACAGCCCATCTCTAATATCCACCACTAACCAAGCCCATCCCGAGCTTATACCTC 1157

QY 235 ACCTACCCCAACCTCCCGAGGGAGCAGCTATTTAAGGGAGCAGAGTGCAGAACAA 294
    |||||
Db 1156 ACCTACCCCAACCTCCCGAGGGAGCAGCTATTTAAGGGAGCAGAGTGCAGAACAA 1097
```

```
Qy 295 CAAGACGGCTGGGATACAACTCTGGAGTCCTCTGA 331
| | | | | | | | | | | | | | | | | | | | | |
Db 1096 CAAAACGACCTAAAAATACAACTCTAAAAATCCTCTAA 1060

RESULT 10
LOCUS AX344469/c AX344469 5728 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 316 from Patent WO0200926.
ACCESSION AX344469
VERSION AX344469.1 GI:18492357
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM synthetic construct
          artificial sequences.
REFERENCE
1.
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with signal transduction
JOURNAL
EpiGenomics AG (DE)
FEATURES
source
1. .5728
   /organism="synthetic construct"
   /db_xref="taxon:32630"
   /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 1462 a 79 c 1271 g 2916 t
ORIGIN

Query Match 49.7%; Score 165; DB 6; Length 5728;
Best Local Similarity 74.7%; Pred. No. 1.5e-42;
Matches 207; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 55 ATTCCAGCGCGGCAAACTGCTGAAATGTGTTGGCATCAGCTACTGACACGTAAGGT 114
| | | | | | | | | | | | | | | | | | | | | |
Db 981 ATAAAAAACTACAAACTACTTAAATATATTTTAACTCACTACTACAGTAAAT 922

Qy 115 TTCCCAATCCTCACTGTCTCTCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGG 174
| | | | | | | | | | | | | | | | | | | | | |
Db 921 TTCCCAATCCTCACTGTCTCTCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGG 174

Qy 175 GTATGGGACCAATCTCTGAGTCCACCACTGACACGCCATCCCGCTTGTGCTC 234
| | | | | | | | | | | | | | | | | | | | | |
Db 861 ATATAACGACCAATCTCTGAGTCCACCACTGACACGCCATCCCGCTTGTGCTC 234

Qy 235 ACCTACCCCAACTCCCGAGGAGGAGCTATTAAAGGGGAGGAGGATTCAGAACAAA 294
| | | | | | | | | | | | | | | | | | | | | |
Db 801 ACTACCCCAACTCCCGAGGAGGAGCTATTAAAGGGGAGGAGGATTCAGAACAAA 294

Qy 295 CAAGACGGCTGGGATACAACTCTGGAGTCCTCTGA 331
| | | | | | | | | | | | | | | | | | | | | |
Db 741 CAAAACGACCTAAAAATACAACTCTAAAAATCCTCTAA 705

RESULT 11
AX251472/c AX251472 6083 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 440 from Patent WO0168912.
ACCESSION AX251472
VERSION AX251472.1 GI:15984895
KEYWORDS synthetic construct.
SOURCE synthetic construct
          artificial sequences.
ORGANISM synthetic construct
          1 (bases 1 to 6083)
REFERENCE
1.
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
          oncogenes
JOURNAL Patent: WO 0168912-A 440 20-SEP-2001;
EpiGenomics AG (DE)
FEATURES Location/Qualifiers
source
1. .6083
   /organism="synthetic construct"
   /db_xref="taxon:32630"
```

```
BASE COUNT 1525 a 83 c 1367 g 3108 t
ORIGIN

Query Match 49.7%; Score 165; DB 6; Length 6083;
Best Local Similarity 74.7%; Pred. No. 1.5e-42;
Matches 207; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 55 ATTCCAGCGCGGCAAACTGCTGAAATGTGTTGGCATCAGCTACTGACACGTAAGGT 114
| | | | | | | | | | | | | | | | | | | | | |
Db 1336 ATAAAAAACTACAAACTACTTAAATATATTTTAACTCACTACTACAGTAAAT 1277

Qy 115 TTCCCAATCCTCACTGTCTCTCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGG 174
| | | | | | | | | | | | | | | | | | | | | |
Db 1276 TTCCCAATCCTCACTGTCTCTCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGG 174

Qy 175 GTATGGGACCAATCTCTGAGTCCACCACTGACACGCCATCCCGCTTGTGCTC 234
| | | | | | | | | | | | | | | | | | | | | |
Db 1216 ATATAACGACCAATCTCTGAGTCCACCACTGACACGCCATCCCGCTTGTGCTC 234

Qy 235 ACCTACCCCAACTCCCGAGGAGGAGCTATTAAAGGGGAGGAGGATTCAGAACAAA 294
| | | | | | | | | | | | | | | | | | | | | |
Db 1156 ACCTACCCCAACTCCCGAGGAGGAGCTATTAAAGGGGAGGAGGATTCAGAACAAA 294

Qy 295 CAAGACGGCTGGGATACAACTCTGGAGTCCTCTGA 331
| | | | | | | | | | | | | | | | | | | | | |
Db 1096 CAAAACGACCTAAAAATACAACTCTAAAAATCCTCTAA 1060

RESULT 12
AX348873 AX348873 5356 bp DNA linear PAT 06-FEB-2002
LOCUS Sequence 331 from Patent WO0202807.
DEFINITION
ACCESSION AX348873
VERSION AX348873.1 GI:18614908
KEYWORDS synthetic construct.
SOURCE synthetic construct
          artificial sequences.
ORGANISM
REFERENCE
1.
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with cell signalling
JOURNAL Patent: WO 0202807-A 331 10-JAN-2002;
EpiGenomics AG (DE)
FEATURES Location/Qualifiers
source
1. .5356
   /organism="synthetic construct"
   /db_xref="taxon:32630"
   /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 1476 a 60 c 1356 g 2464 t
ORIGIN

Query Match 45.2%; Score 150.2; DB 6; Length 5356;
Best Local Similarity 72.7%; Pred. No. 1e-37;
Matches 194; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 66 CCAAACTGCTGAAATGTCTTTGGCATCAGCTACTGACACGTAAGGTTCCTCAATCCT 125
| | | | | | | | | | | | | | | | | | | | | |
Db 4032 GTAAATTTGTTGAAATGTCTTTGGTATTGATTATGATGATGATGATGATGATGAT 4091

Qy 126 CAACTCTGCTGCCAGCTGATGAGGGGAGGAAAGGATTAACCTAGGGGTATGGCGGAC 185
| | | | | | | | | | | | | | | | | | | | | |
Db 4092 TAATTTGTTTGTGTTAGTTAGTGGGAGGAAAGGATTAATTTAGGGGTATGGCGGAT 4151

Qy 186 CAATCTGAGTCCACCACTGACCAAGCCATCCCGCTTGTGCTCCTCCTACCTACCCCA 245
| | | | | | | | | | | | | | | | | | | | | |
Db 4152 TAAATTTGAGTTTATTAATTTGATGATGATGATGATGATGATGATGATGATGAT 4211

Qy 246 ACCTCCCAAGGGGAGGAGCTATTAAAGGGGAGGAGGAGGATTAACCTAGGGGTATGGCGGAC 305
| | | | | | | | | | | | | | | | | | | | | |
Db 4212 ATTTTATGAGGGGAGTATGATTTTAAAGGGGAGGAGGAGGATTAACCTAGGGGTAT 4271

Qy 306 GGGGATACAACTCTGGAGTCCTCTGAG 332
```

```
FEATURES
  source
    Location/Qualifiers
      EpiGenomics AG (DE)
      1..6083
      /organism="synthetic construct"
      /db_xref="taxon:32630"
      /note="chemically treated genome"
```

FEATURES	SOURCE
basecat	1_997
location/Qualifiers	
organism="Homo sapiens"	
db_xref="LOCUSID:5320"	
db_xref="taxon:9606"	
clone="MGC:14516 IMAGE:4274550"	
tissue_type="Prostate"	
clone_type="NTH MGC 83"	

CDS  
/lab\_host="DH103"  
/note="Vector: pDNR-LIB"  
273...707  
/codon\_start=1  
/product="phospholipase A2, group IIA (platelets, synovial fluid)"  
/protein\_id="AAH05919.1"  
/db\_xref="GI:13543521"  
/translation="MKTLILLAVIMIFGLQAHGNLVNFRHMIKLTGKEAALSYGFY  
GCHGVGGRSPADTDRCVTHDCYKRLKRGCGTKFLSYKFSNSGSRITCAKQDS  
CRSQLCECDKAAATCFARNKTTYNKKYQYISNKHCRGSTPRC"  
BASE COUNT 289 a 279 c 231 g 198 t  
ORIGIN

Query Match 35.8%; Score 118.8; DB 9; Length 997;  
Best Local Similarity 83.3%; Pred. No. 1.8e-27;  
Matches 135; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
Qy 171 AGGGTATGGCGACCAATCTGAGTCCACCACTGACACGCGCCATCCCGACCTTGTG 230  
Db 3 AGGAAAAGAGCAACAGATCCAGGGAGCATTCACCTGCCCTGTCTCCAAACAGCCTTGTG 62  
Qy 231 CCTCACCTACCCCAACTCCAGAGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGAA 290  
Db 63 CCTCACCTACCCCAACTCCAGAGGAGCAGCTATTTAAGGGAGCAGGAGTGCAGAA 122  
Qy 291 CAAACAAGACGGCCTGGGGATACAACTCTGGAGTCCTCTGAG 332  
Db 123 CAAACAAGACGGCCTGGGGATACAACTCTGGAGTCCTCTGAG 164

Search completed: February 19, 2003, 23:10:05  
Job time : 1140.56 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:26:38 ; Search time 201.781 Seconds  
(without alignments)  
3705.323 Million cell updates/sec

Title: US-09-808-388-6  
Perfect score: 332  
Sequence: 1 GTaccatttcgacaaacta.....caactctggagtcctctgag 332

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :	N_Geneseq_101002.*
1:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332	100.0	332	22	Partial synthetic
2	271	81.6	271	22	Partial human PLA2
3	268.4	80.8	6172	10	HindIII fragment o
4	268.2	80.8	1116	20	Human normal ovari
5	213	64.2	1080	10	Nucleotide sequenc
6	165	49.7	5356	24	Chemically treated
7	165	49.7	5728	24	Signal transductio
8	165	49.7	6083	22	Tumour suppressor
9	150.2	45.2	5356	24	Chemically treated

10	150.2	45.2	5728	24	ABK31472	Signal transductio
11	150.2	45.2	6083	22	AAS46714	Tumour suppressor
12	118.8	35.8	204	23	ABV03174	Human prostate exp
13	118.8	35.8	481	23	ABV39325	Human prostate exp
14	84.4	25.4	3330	24	ABK63588	Rat sequence diffe
15	80.8	24.3	1076	21	AAF15635	Human prostate can
16	79.2	23.9	1160	23	ABV22259	Human prostate exp
17	79.2	23.9	1160	23	ABV23366	Human prostate exp
18	79.2	23.9	1160	23	ABV23388	Human prostate exp
19	79.2	23.9	1160	23	ABV25535	Human prostate exp
20	79.2	23.9	1160	23	ABV28096	Human prostate exp
21	79.2	23.9	1160	23	ABV29221	Human prostate exp
22	79.2	23.9	1160	23	ABV29244	Human prostate exp
23	78.4	23.6	126	23	ABV14604	Human prostate exp
24	78.4	23.6	165	23	ABV05435	Human prostate exp
25	78.4	23.6	516	23	ABV35682	Human prostate exp
26	78.4	23.6	516	23	ABV44487	Human prostate exp
27	48.2	14.5	119	23	ABV08557	Human prostate exp
28	48.2	14.5	390	23	ABV38459	Human prostate exp
29	41	12.3	41	22	AAI64305	PPAR response elem
30	37.6	11.3	441	23	ABV38462	Human prostate exp
31	37	11.1	83	23	ABV14206	Human prostate exp
32	37	11.1	119	23	ABV08718	Human prostate exp
33	37	11.1	120	23	ABV08560	Human prostate exp
34	37	11.1	276	23	ABV08936	Human prostate exp
35	37	11.1	387	23	ABV44134	Human prostate exp
36	37	11.1	428	23	ABV35300	Human prostate exp
37	37	11.1	448	23	ABV38616	Human prostate exp
38	37	11.1	450	23	ABV38824	Human prostate exp
39	36	10.8	123	23	ABV05037	Human prostate exp
40	34	10.2	1687	23	ABL12468	Drosophila melanog
41	34	10.2	1687	23	ABL18110	Drosophila melanog
42	34	10.2	1687	23	ABL18210	Drosophila melanog
43	33.8	10.2	420	21	AAC61755	cDNA encoding a hu
44	33.8	10.2	1441	21	AAC61753	cDNA encoding a hu
45	33.8	10.2	2236	21	AAC61749	cDNA encoding a hu

ALIGNMENTS

RESULT 1  
AAI64308  
ID AAI64308 standard; DNA; 332 BP.  
AC AAI64308;  
DT 15-NOV-2001 (first entry)  
DE Partial synthetic PLA2SIIA gene promoter.  
XX PPAR response element; antiinflammatory; antithrombotic; cytostatic;  
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2SIIA;  
KW peroxisome proliferator activated receptor;  
KW secreted non-pancreatic phospholipase A2; ds.  
OS Synthetic.  
XX WO200168845-A2.  
FN 20-SEP-2001.  
PD 14-MAR-2001; 2001WO-FR00759.  
PF 14-MAR-2001; 2000FR-0003262.  
PR 13-APR-2000; 2000US-0196959.  
XX (AVET ) AVENTIS PHARMA SA.  
PI Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;  
DR WPI; 2001-582451/65.  
XX

PT New hybrid promoter induced by inflammation, useful in gene therapy of  
PT arthritis, comprises peroxisome proliferator activated receptor  
XX response element and promoter of secreted phospholipase A2 -  
XX Disclosure; Page 51-52; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR  
CC (peroxisome proliferator activated receptor) response element (PPRE); and  
CC (ii) at least part of the promoter of the PLA2sIIA (secreted  
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to  
CC regulate expression of therapeutic transgenes, for experimental,  
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for  
CC treatment of arthritis, but also in bone, muscle, liver, heart, the  
CC nervous system and tumours. The present sequence is a partial synthetic  
CC PLA2sIIA promoter sequence, which was used to generate the hybrid  
CC promoter of the present invention.

XX Sequence 332 BP; 96 A; 91 C; 82 G; 63 T; 0 other;  
SQ Query Match 100.0%; Score 332; DB 22; Length 332;  
Best Local Similarity 100.0%; Pred. No. 7.1e-98;  
Matches 332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTACCAATTCGACAACTAGTCAAGGTCATCAAACTAGTCAAGGTCAAATTGCA 60  
Db 1 GTACCAATTCGACAACTAGTCAAGGTCATCAAACTAGTCAAGGTCAAATTGCA 60

Qy 61 ACGGGGCAAACTGCTTGAATGTTTGGCATCAGTACTGACACGTAAGGTTTCCCA 120  
Db 61 ACGGGGCAAACTGCTTGAATGTTTGGCATCAGTACTGACACGTAAGGTTTCCCA 120

Qy 121 ATCTCAACTCTGTCCTGCCAGCTGATGAGGGGAAGGAAGGATTACCTAGGGGTATGG 180  
Db 121 ATCTCAACTCTGTCCTGCCAGCTGATGAGGGGAAGGAAGGATTACCTAGGGGTATGG 180

Qy 181 GCGACCAATCTGAGTCCACCACTGACACGCGCCATCCCGAGCTTGTGCTTCACTAC 240  
Db 181 GCGACCAATCTGAGTCCACCACTGACACGCGCCATCCCGAGCTTGTGCTTCACTAC 240

Qy 241 CCCCAACCTCCAGAGGAGCAGCTATTAAAGGGAGCAGGATGAGAACCAACAGAC 300  
Db 241 CCCCAACCTCCAGAGGAGCAGCTATTAAAGGGAGCAGGATGAGAACCAACAGAC 300

Qy 301 GGCCTGGGGATACAACTCTGGAGTCCTCTGAG 332  
Db 301 GGCCTGGGGATACAACTCTGGAGTCCTCTGAG 332

RESULT 2  
AAI64307  
ID AAI64307 standard; DNA; 271 BP.  
AC AAI64307;  
DT 15-NOV-2001 (first entry)  
XX Partial human PLA2sIIA gene promoter.  
DE PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2sIIA;  
XX peroxisome proliferator activated receptor; human;  
KW secreted non-pancreatic phospholipase A2; ds.  
XX Homo sapiens.  
XX OS  
XX WO200168845-A2.  
XX PN  
XX PD 20-SEP-2001.  
XX PF 14-MAR-2001; 2001WO-FR00759.  
XX PR 14-MAR-2000; 2000FR-0003262.  
XX PR 13-APR-2000; 2000US-0196959.

XX (AVET ) AVENTIS PHARMA SA.  
PA Massaad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;  
XX WI; 2001-582451/65.  
XX New hybrid promoter induced by inflammation, useful in gene therapy of  
XX arthritis, comprises peroxisome proliferator activated receptor  
PT response element and promoter of secreted phospholipase A2 -  
XX Claim 5; Page 51; 52pp; French.

XX The present invention relates to a hybrid promoter comprising (i) a PPAR  
CC (peroxisome proliferator activated receptor) response element (PPRE); and  
CC (ii) at least part of the promoter of the PLA2sIIA (secreted  
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to  
CC regulate expression of therapeutic transgenes, for experimental,  
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for  
CC treatment of arthritis, but also in bone, muscle, liver, heart, the  
CC nervous system and tumours. The present sequence is a partial human  
CC PLA2sIIA promoter sequence, which was used to generate the hybrid  
CC promoter of the present invention.

XX Sequence 271 BP; 70 A; 79 C; 71 G; 51 T; 0 other;  
SQ Query Match 81.6%; Score 271; DB 22; Length 271;  
Best Local Similarity 100.0%; Pred. No. 4.5e-78;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 CGCGGCAAACTGCTTGAATGTTTGGCATCAGTACTGACACGTAAGGTTTCCCA 121  
Db 1 CGCGGCAAACTGCTTGAATGTTTGGCATCAGTACTGACACGTAAGGTTTCCCA 60

Qy 122 TCCTCAACTCTGTCCTGCCAGCTGATGAGGGGAAGGAAGGATTACCTAGGGGTATGG 181  
Db 61 TCCTCAACTCTGTCCTGCCAGCTGATGAGGGGAAGGAAGGATTACCTAGGGGTATGG 120

Qy 182 CGACCAATCTGAGTCCACCACTGACACGCGCCATCCCGAGCTTGTGCTTCACTACC 241  
Db 121 CGACCAATCTGAGTCCACCACTGACACGCGCCATCCCGAGCTTGTGCTTCACTACC 180

Qy 242 CCCCAACCTCCAGAGGAGCAGCTATTAAAGGGAGCAGGATGAGAACCAACAGAC 301  
Db 181 CCCCAACCTCCAGAGGAGCAGCTATTAAAGGGAGCAGGATGAGAACCAACAGAC 240

Qy 302 GGCCTGGGGATACAACTCTGGAGTCCTCTGAG 332  
Db 241 GGCCTGGGGATACAACTCTGGAGTCCTCTGAG 271

RESULT 3  
AAI91825  
ID AAI91825 standard; DNA; 6172 BP.  
XX AC AAI91825;  
XX 31-JUL-1992 (second entry)  
XX HindIII fragment of PLA2 8.5 EMBL3 encoding human inflammatory  
DE phospholipase A2.  
XX Inflammation; acid stable; phosphatide 2-acylhydrolase; lipolytic;  
XX glycerophospholipids; non-pancreatic; ss.  
XX Homo sapiens.  
XX OS  
XX FH Key Location/Qualifiers  
XX exon 1..2492 a  
XX FT /\*tag= a  
XX FT /\*tag= 1  
XX FT 2702..2846  
XX FT /\*tag= b



103 T /number= 2  
104 T 3105...3211  
105 T /tag= c  
106 T /number= 3  
107 T 5383...6172  
108 T /tag= d  
109 T /number= 4  
110 T 2453...2721  
111 T /tag= e  
112 T sig\_peptide 5771...5776  
113 T /tag= f  
114 T misc\_feature 2715...2826  
115 T /tag= g  
116 T /note="Claim 24"  
117 T 2715...2826  
118 T /tag= g  
119 T /note="Claim 24"  
120 X  
121 X WO8909818-A.  
122 X  
123 X 19-OCT-1989.  
124 X  
125 X 11-APR-1989; 89WO-US01418.  
126 X  
127 X 15-APR-1988; 88US-0181893.  
128 X  
129 X (BIOJ ) BIOGEN INC.  
130 X  
131 X Kramer RM, Pepinsky RB, Hession C;  
132 X  
133 X WPI; 1989-324225/44.  
134 X P-PSDB; RAP93112.  
135 X  
136 X Acid stable phospholipase A2 - used for prodn. of antibodies and in  
137 X the treatment or diagnosis of inflammation of diseases.  
138 X  
139 X Claim 27; Fig 12; 84pp; English.  
140 X  
141 X A genomic DNA library was prep'd. from a mutant fibroblast cell line  
142 X which contains 5 copies of the X chromosome (GM5009). The plaques  
143 X were screened for a gene encoding PLA2 using probes designed from  
144 X peptides derived from the purified protein. A positive clone, PLA2  
145 X 8.5 EMBL3 was purified and a 6.2 kb insert sequenced (shown here).  
146 X Corresponding cDNA sequences (i.e. without introns; see N97209) can  
147 X be ligated into expression vectors for the prodn. of recombinant  
148 X PLA2. The protein, and antibodies raised to it, can be used for  
149 X diagnosis of inflammation and tissue injury associated with various  
150 X diseases.  
151 X See also N91826-33 and N97209.  
152 X  
153 X Sequence 6172 BP; 1624 A; 1408 C; 1740 G; 1400 T; 0 other;  
154 X  
155 X Query Match 80.8%; Score 268.4; DB 10; Length 6172;  
156 X Best Local Similarity 97.8%; Pred. No. 1.1e-75;  
157 X Matches 272; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
158 X  
159 QY 55 ATTGGAAGCGGCAAAAGTGGTGAATGTGTTTGGCATCAGTACTGACGTAAGGT 114  
160 DB 1369 ATGGAAGAGTCAAAAGTGGTGAATGTGTTTGGCATCAGTACTGACGTAAGGT 1428  
161 QY 115 TTCCCAATCTCAACTCTGCTCCGACGCTGATGAGGGGAGGAAGGATTACCTAGGG 174  
162 DB 1429 TTCCCAATCTCAACTCTGCTCCGACGCTGATGAGGGGAGGAAGGATTACCTAGGG 1488  
163 QY 175 GTATGGCGGCAACCAATCTCGAGTCCACCACTGACACGCCCATCCAGCGTTGTGCTC 234  
164 DB 1489 GTATGGCGGCAACCAATCTCGAGTCCACCACTGACACGCCCATCCAGCGTTGTGCTC 1548  
165 QY 235 ACTTACCCCACTCCAGAGGAGGAGCTATTATTAAGGGAGCAGGAGTGCAGACAA 294  
166 DB 1549 ACTTACCCCACTCCAGAGGAGGAGCTATTATTAAGGGAGCAGGAGTGCAGACAA 1608  
167 QY 295 CAAGACGGGCTGGGATACAACTCTGGAGTCTCTGAG 332

103 DB 1609 CAAGACGGGCTGGGATACAACTCTGGAGTCTCTGAG 1646  
104  
105 RESULT 4  
106 AAZ41274  
107 ID AAZ41274 standard; cDNA; 1116 BP.  
108 XX AAZ41274;  
109 AC AAZ41274;  
110 DT 18-JAN-2000 (first entry)  
111 XX Human normal ovarian tissue derived cDNA 53.  
112 DE Human; ovary; screening; ovarian cancer; treatment; ss.  
113 XX Homo sapiens.  
114 OS  
115 PN D319816395-A1.  
116 XX  
117 PD 07-OCT-1999.  
118 XX  
119 PF 03-APR-1998; 98DE-1016395.  
120 XX  
121 PR 03-APR-1998; 98DE-1016395.  
122 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
123 PA  
124 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
125 XX WPI; 1999-552352/47.  
126 XX  
127 PT Nucleic acid sequences potentially useful in diagnosis or therapy of  
128 PT ovarian cancer -  
129 XX  
130 PS Claim 3; Page 164; 274ppp; German.  
131 CC  
132 CC This invention describes novel nucleic acid sequences that are highly  
133 CC expressed in normal ovary tissue. Artificial chromosomes and cosmid  
134 CC clones containing the sequences can be used as gene transfer vehicles.  
135 CC The sequences can be used to produce DNA fragments containing  
136 CC full-length genes. Host cells transformed with the sequences can be used  
137 CC to produce polypeptides or polypeptide fragments, which can be used to  
138 CC screen phage displays for polypeptides that bind to them, or as tools for  
139 CC identifying agents active against ovarian cancer, or to prepare  
140 CC medicaments for treating ovarian cancer. The cDNA sequences can be used  
141 CC to obtain genomic genes, their promoters, enhancers, silencers, exon  
142 CC structures, intron structures and their splice variants. AAZ41222-241324  
143 CC represent cDNA sequences derived from normal human ovarian tissue and  
144 CC which encode the protein fragments represented in AAY59724-Y59837.  
145 CC  
146 XX SQ Sequence 1116 BP; 311 A; 311 C; 266 G; 228 T; 0 other;  
147  
148 X Query Match 80.8%; Score 268.2; DB 20; Length 1116;  
149 X Best Local Similarity 98.9%; Pred. No. 6.4e-77;  
150 X Matches 270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
151 X  
152 QY 60 AACGGGCAAACTGCTGAAATGTGTTTGGCATCAGTACTGACGTAAGGTTTCCC 119  
153 DB 2 AAGACTGCAAACTGCTGAAATGTGTTTGGCATCAGTACTGACGTAAGGTTTCCC 61  
154 QY 120 AATCTCTCACTCTGCTGCGCAGCTGATGAGGGGAGGAAGGATTACCTAGGGGTATG 179  
155 DB 62 AATCTCTCACTCTGCTGCGCAGCTGATGAGGGGAGGAAGGATTACCTAGGGGTATG 121  
156 QY 180 GCGGACCAATCTGAGTTCACCACTGACACGCCCATCCAGCGTTGTGCTCACCTA 239  
157 DB 122 GCGGACCAATCTGAGTTCACCACTGACACGCCCATCCAGCGTTGTGCTCACCTA 181  
158 QY 240 CCCCCAATCTCCAGAGGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAAGA 299  
159 DB 182 CCCCCAATCTCCAGAGGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAAGA 241

QY 300 CGGCTGGGATACAACTCTGGAGTCTCTCTGAG 332  
 |||||  
 Db 242 CGGCTGGGATACAACTCTGGAGTCTCTCTGAG 274  
 |||||

RESULT 5  
 ID AAN91259 standard; DNA; 1080 BP.  
 XX AAN91259;  
 AC AAN91259;  
 XX 27-JUN-1980 (first entry)  
 DT  
 XX Nucleotide sequence of exon 1 from genomic clone lambda sPLA2-6 of human  
 DE synovial phospholipase 2 (sPLA2) type A.  
 XX Human synovial phospholipase A2 gene; clone lambda sPLA2-6; exon 1.  
 XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH CAAT\_signal 889..893  
 FT /\*tag= a  
 FT TATA\_signal 968..974  
 FT /\*tag= b  
 FT exon 1016..1035  
 FT /\*tag= c  
 FT /\*note="Exon 1"  
 PN WO8901773-A.  
 XX  
 XX 09-MAR-1989.  
 PD  
 XX 23-AUG-1988; 88WO-US02896.  
 PF  
 XX 16-AUG-1988; 88US-0231865, US-089883.  
 PR  
 XX (BIOT-) BIOTECHN RES PARTN (UTOR).  
 PA  
 XX Johnson UK, Seilhamer JU, Pruzanski W, Vada P;  
 PI WPI; 1989-085394/11.  
 XX  
 XX Mammalian synovial phospholipase A2- used in food processing  
 PT design and screening of inflammation inhibitors, as an anticancer  
 PT drug or vaccine adjuvant etc  
 XX  
 XX Fig 7; ; 70pp; English.  
 XX  
 XX EMBL3-human leucocyte genomic library was screened using labelled probes  
 CC (n50885 and n90887) based on sPLA2. Clone lambda sPLA2-6 is one of the  
 CC two unique sPLA2 clones thus identified. sPLA2-6 exons were identified  
 CC using the cDNA sequence in lambda sPLA2cDNA-4 (n91258). There are five  
 CC exons in lambda sPLA2-6. This is the first one. The other four are in  
 CC n91260.  
 XX  
 XX Sequence 1080 BP; 306 A; 242 C; 303 G; 229 T; 0 other;  
 SQ

Query Match 64.28; Score 213; DB 10; Length 1080;  
 Best Local Similarity 93.18; Pred. No. 5.6e-59;  
 Matches 256; Conservative 0; Mismatches 15; Indels 4; Gaps 3;

QY 61 AGCGGCAAACTGCTGAAATGTTTGGCATCTGACAGTAAAG-TTTCCC 119  
 |||||  
 Db 762 ACTCGGCAAACTGCTGAAATGTTTGGCATCTGACAGTAAAGTTCCTCC 821  
 |||||

QY 120 ATCTCTCACTGTCCTG--CCAGCTGATGAGGGAAGGATTAACCTAGGGTA 177  
 |||||  
 Db 822 AATCTCACTGTCCTGCGGCGCTGATGAGGGAAGGATTAACCTAGGGTA 881  
 |||||

QY 178 TGGGGCAACATCTTGAGTCCACCACTGACACGCCATCCCCAGCTTGTGCTCACC 237  
 |||||  
 Db 882 TGGGGCAACATCTTGAGTCCACCACTGACACGCCATCCCCAGCTTGTGCTCACC 941  
 |||||

QY 238 TACCCCAACCTCCAGAGGAGCAGCTATTATTAAGGGAGCAGAGTGCAACAACAA 297  
 |||||  
 Db 942 TACCCCAACCT-CCAGAGGAGCAGCTATTATTAAGGGAGCAGAGTGCAACAACAA 1000  
 |||||

QY 298 GAGGCTGGGATACAACTCTGGAGTCTCTCTGAG 332  
 |||||  
 Db 1001 GAGGCTGGGATACAACTCTGGAGTCTCTCTGAG 1035  
 |||||

RESULT 6  
 ID ABL70442/c  
 XX ABL70442 standard; DNA; 5356 BP.  
 AC ABL70442;  
 XX 01-JUL-2002 (first entry)  
 DT  
 XX Chemically treated cell signalling DNA sequence complementary to#166.  
 DE  
 XX Cell signalling; cytosine methylation; cell signalling disease;  
 KW cancer; tumour; cytostatic; ds.  
 XX  
 XX Unidentified.  
 OS  
 XX WO200202807-A2.  
 PN  
 XX 10-JAN-2002.  
 PD  
 XX 29-JUN-2001; 2001WO-EP07471.  
 PF  
 XX 30-JUN-2000; 2000DE-1032529.  
 PR  
 XX 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI WPI; 2002-154758/20.  
 XX  
 XX Nucleic acid, useful for diagnosis and therapy of diseases associated  
 PT with cell signalling e.g. cancer, comprises chemically modified genomic  
 PT sequences of genes associated with cell signalling -  
 XX  
 XX Claim 1; SEQ ID NO 332; 24pp+sequence listing; English.  
 XX  
 XX The invention relates to a nucleic acid comprising a sequence of at least  
 CC 18 bases of a segment of chemically pretreated DNA of genes associated  
 CC with cell signalling. The activity of the modified sequences of the  
 CC invention may be described as cytostatic. The object of the invention is  
 CC to provide the chemically modified DNA of genes associated with cell  
 CC signalling, as well as oligonucleotides and/or PNA-oligomers for  
 CC detecting cytosine methylations, as well as a method which is  
 CC particularly suitable for the diagnosis and/or therapy of genetic and  
 CC epigenetic parameters of genes associated with cell signalling. The  
 CC chemically modified DNA provided by the invention is useful for diagnosis  
 CC and therapy of diseases such as solid tumours and cancer. The sequences  
 CC given in records ABL70111-ABL70626 represent chemically pre-treated  
 CC genomic DNA's of genes associated with cell signalling.  
 CC Note: The sequence data for this patent is not represented in the printed  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.  
 XX  
 XX Sequence 5356 BP; 1357 A; 60 C; 1167 G; 2772 T; 0 other;  
 SQ

Query Match 49.7%; Score 165; DB 24; Length 5356;  
 Best Local Similarity 74.7%; Pred. No. 4.3e-43;  
 Matches 207; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 55 ATTGACGCGGCAAACTGCTGAAATGTTTGGCATCTGACAGTAAAGT 114  
 |||||  
 Db 1336 ATAAAAAAATACAAAACCTACCTAAATATATTTTAACTACTACACGCTAAAT 1277  
 |||||



CC bisulphite, of genes associated with tumour suppression and  
CC oncogenes having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (Se) and sequences complementary to (Se). The nucleic acid may be a  
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with CpG dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are disadvantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes. Sequences with even numbered seq ID numbers are the  
CC complementary sequence of the corresponding odd numbered sequence (e.g.  
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
CC is missing).  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 6083 BP; 1525 A; 83 C; 1367 G; 3108 T; 0 other;  
Query Match 49.7%; Score 165; DB 22; Length 6083;  
Best Local Similarity 74.7%; Pred. No. 4.5e-43;  
Matches 207; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 55 ATTCAAGCGGCAAACTGCTGAAATGTTTGGCATCAGCTACTGACGTAAGCT 114  
DB 1336 ATAAAAAACTACAAAACCTAACTAAATATATATTTTAACTCACTACACGTAAT 1277  
QY 115 TTCCCAATCCTCAACTCTGCTGCGAGCTGATCAGGGGAAGGAAGGATTACCTAGG 174  
DB 1276 TTCCCAATCCTCAACTCTGCTGCGAGCTGATCAGGGGAAGGAAGGATTACCTAGG 1217  
QY 175 GTATGGGGACCAATCTCTGAGTCACCACTGACGACGCGCATCCCGAGCTGTGCTTC 234  
DB 1216 ATATAAGGACCAATCTCTGAGTCACCACTGACGACGCGCATCCCGAGCTGTGCTTC 1157  
QY 235 ACCTACCCCACTCCAGAGGAGCAGCTATTTAAGGGAGCAGGATGCGAGACAA 294  
DB 1156 ACCTACCCCACTCCAGAGGAGCAGCTATTTAAGGGAGCAGGATGCGAGACAA 1097  
QY 295 CAAGAGCGCTGGGATACAACTCTGGAGTCTCTGA 331  
DB 1096 CAAGAGCGCTGGGATACAACTCTGGAGTCTCTGA 1060

RESULT 9  
ID ABL70441  
ABL70441 standard; DNA; 5356 BP.

XX AC ABL70441;

XX DT 01-JUL-2002 (first entry)

XX DE Chemically treated cell signalling DNA sequence#166.

XX KW Cell signalling; cytosine methylation; cell signalling disease;  
XX KW cancer; tumour; cytostatic; ds.

XX OS Unidentified.

XX FN WO200202807-A2.

XX PD 10-JAN-2002.

XX PF 29-JUN-2001; 2001WO-EP07471.

PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-154758/20.  
XX Nucleic acid, useful for diagnosis and therapy of diseases associated  
XX with cell signalling e.g. cancer, comprises chemically modified genomic  
XX sequences of genes associated with cell signalling.  
XX Claim 1; SEQ ID NO 331; 24pp+sequence listing; English.  
XX The invention relates to a nucleic acid comprising a sequence of at least  
XX 18 bases of a segment of chemically pretreated DNA of genes associated  
XX with cell signalling. The activity of the modified sequences of the  
XX invention may be described as cytostatic. The object of the invention is  
XX to provide the chemically modified DNA of genes associated with cell  
XX signalling, as well as oligonucleotides and/or PNA-oligomers for  
XX detecting cytosine methylations, as well as a method which is  
XX particularly suitable for the diagnosis and/or therapy of genetic and  
XX epigenetic parameters of genes associated with cell signalling. The  
XX chemically modified DNA provided by the invention is useful for diagnosis  
XX and therapy of diseases such as solid tumours and cancer. The sequences  
XX given in records ABL70111-ABL70626 represent chemically pre-treated  
XX genomic DNA's of genes associated with cell signalling.  
XX Note: The sequence data for this patent is not represented in the printed  
XX specification, but is based on sequence information supplied by the  
XX European Patent Office.

XX SQ Sequence 5356 BP; 1476 A; 60 C; 1356 G; 2464 T; 0 other;

Query Match 45.2%; Score 150.2; DB 24; Length 5356;  
Best Local Similarity 72.7%; Pred. No. 2.8e-38;  
Matches 194; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 66 GCAAACTGCTGAATGTTTGGCATCAGCTACTGACGTAAGGTTTCCCAATCCT 125

DB 4032 GTAAATGTTGAATGTTTGGTATTGTTATTGATGTAAGGTTTAAATTT 4091

QY 126 CAATCTGTCTGCGAGCTGATGAGGGGAAGGAAGGATTAAGGGGATGGCGAC 185

DB 4092 TAATTTGTTTGTAGTTGATGAGGGGAAGGAAGGATTAATTTAGGGGATGGCGAT 4151

QY 186 CAATCTGAGTCCACCACTGACGCGCCATCCCGAGCTTGCTCCTACCTACCCCA 245

DB 4152 TAATTTGAGTTTAAATGATTAAGGAGTATTTAGTTTGTGTTTATTTT 4211

QY 246 ACCTCCGAGGGGAGCAGCTATTTAAGGGGAGCAGGATGCGAGAACAAACAGCGGCT 305

DB 4212 ATTTTGTAGAGGAGTAGTTTATTTAAGGGGAGTAGGAGTAGAATAAATAAGCGGTT 4271

QY 306 GGGATACAACTCTGGAGTCTCTGAG 332

DB 4272 GGGATATAATTTGGAGTTTGTAG 4298

RESULT 10

ABK31472

ID ABK31472 standard; DNA; 5728 BP.

XX AC ABK31472;

XX DT 23-APR-2002 (first entry)

XX DE Signal transduction associated gene modified DNA #158.

XX KW Human; signal transduction associated gene; cytosine methylation state;  
XX KW CpG island; signal transduction associated disease; solid tumour; cancer;  
XX KW antitumour; cytostatic; mutant; ds.

JS Homo sapiens.  
 JS Synthetic.  
 XX WO200200926-A2.  
 XX 03-JAN-2002.  
 XX 29-JUN-2001; 2001WO-EP07472.  
 XX 30-JUN-2000; 2000DE-1032529.  
 XX 01-SEP-2000; 2000DE-1043926.  
 XX (EPIG-) EPIGENOMICS AG.  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2002-147896/19.  
 XX Oligonucleotide for diagnosis and therapy of diseases associated with  
 XX signal transduction e.g. cancer, comprises chemically modified genomic  
 XX sequences of genes associated with signal transduction -  
 XX Claim 1; SEQ ID No 315; 24pp; English.  
 XX The present invention relates to chemically modified DNA sequences of  
 XX signal transduction associated genes. The DNA sequences are chemically  
 XX modified using a solution of bisulphite, hydrogen sulphite or  
 XX disulphite. Also disclosed are oligonucleotides and/or PNA oligomers  
 XX for detecting the cytosine methylation state (CpG islands) of these  
 XX genes, and a method for the diagnosis and/or therapy of genetic and  
 XX epigenetic parameters of genes associated with signal transduction.  
 XX The genomic DNA can be obtained from cells or cellular components which  
 XX contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,  
 XX cerebral-spinal fluid, tissue embedded in paraffin such as tissue from  
 XX eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,  
 XX histologic object slides, and all their possible combinations. The  
 XX sequences of the invention are useful for the diagnosis and therapy of  
 XX diseases associated with signal transduction e.g. solid tumours and  
 XX cancer. ABK3158-ABK31545 represent chemically pretreated genomic DNA  
 XX sequences of different genes associated with signal transduction, or  
 XX their complementary sequences.  
 XX Note: The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from the  
 XX European Patent Office.  
 XX Sequence 5728 BP; 1547 A; 79 C; 1448 G; 2654 T; 0 other;  
 XX  
 XX Query Match 45.2%; Score 150.2; DB 24; Length 5728;  
 XX Best Local Similarity 72.7%; Pred. No. 2.9e-38;  
 XX Matches 194; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
 QY 66 GCAAACTCCTGAAATGTTTGGCATCAGTACTGACACGTAAAGTTTCCCAATCCT 125  
 Db 4759 GTAAATTTGTTTGAATGTTTGGTATTATTGTTTATGATACGTAAAGTTTAAATTT 4818  
 QY 126 CAACTCTGCTGCCAGCTGATGAGGGGAAGAAAGGATTAACCTAGGGTATGGGGAC 185  
 Db 4819 TAATTTGTTTGTAGTTGATGAGGGGAGGAGGATTAATAGGGTATGGGGAT 4878  
 QY 186 CAATCTGATGTCACCACTGACACAGGCCATCCCGACCTTGGCTCCTACCTACCCCA 245  
 Db 4879 TAATTTGATTTTAAATGATTGATTAATTTTATTTTATTTTATTTTATTTT 4938  
 QY 246 ACCTCCAGAGGAGCAGCTATTAGGGGAGCAGGATGCGAGCAACAAAGACGGCCT 305  
 Db 4939 ATTTTATAGAGGAGTAGTTTATTTAAGGGAGTAGGATGTAGATAATTAAGCGTTT 4998  
 QY 306 GGGGATACAACTCTGAGTCCTCTGAG 332  
 Db 4999 GGGGATATATTTTGGAGTTTGTGAG 5025

RESULT 11

AAS46714  
 ID AAS46714 standard; DNA; 6083 BP.  
 XX  
 AC AAS46714;  
 XX  
 XX 18-DEC-2001 (first entry)  
 XX Tumour suppressor gene derived chemically modified sequence #437.  
 XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
 XX cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 XX cytosine methylation; ds.  
 XX Homo sapiens.  
 XX WO200168912-A2.  
 XX 20-SEP-2001.  
 XX 15-MAR-2001; 2001WO-EP02955.  
 XX 15-MAR-2000; 2000DE-1013947.  
 XX 06-APR-2000; 2000DE-1019058.  
 XX 07-APR-2000; 2000DE-1019173.  
 XX 30-JUN-2000; 2000DE-1032529.  
 XX 01-SEP-2000; 2000DE-1043926.  
 XX (EPIG-) EPIGENOMICS AG.  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2001-602752/69.  
 XX Fragments of chemically modified genes associated with tumour suppressor  
 XX genes and oncogenes, useful in designing primers and probes for  
 XX analysing diseases associated with cytosine methylation state e.g.  
 XX cancer -  
 XX Claim 1; SEQ ID No 437; 27pp; English.  
 XX The invention relates to a nucleic acid comprising a sequence of 18  
 XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 XX bisulphite, of genes associated with tumour suppression and  
 XX oncogenes having a sequence taken from 536 (actually 533 since  
 XX numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 XX (S) and sequences complementary to (Ss). The nucleic acid may be a  
 XX peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
 XX form part of a set of probes for detecting the cytosine methylation state  
 XX and/or single nucleotide polymorphisms and also to be used in an  
 XX array for analysing diseases associated with CpG dinucleotides e.g.  
 XX cancers and tumours. The probes can also be used in a method for  
 XX ascertaining genetic and/or epigenetic parameters for the diagnosis  
 XX and/or therapy of existing diseases or the predisposition to specific  
 XX diseases, by analysing cytosine methylations. The parameters may be  
 XX compared to another set of genetic and/or epigenetic parameters, the  
 XX differences serving as basis for diagnosis and/or prognosis events which  
 XX are disadvantageous to patients. The present sequence is one of the  
 XX 533 genomic sequences derived from tumour suppressor genes and  
 XX oncogenes.  
 XX Note: The sequence data for this patent did not form part  
 XX of the printed specification, but was obtained in electronic  
 XX format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 6083 BP; 1640 A; 83 C; 1551 G; 2809 T; 0 other;  
 XX  
 XX Query Match 45.2%; Score 150.2; DB 22; Length 6083;  
 XX Best Local Similarity 72.7%; Pred. No. 2.9e-38;  
 XX Matches 194; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
 QY 66 GCAAACTCCTGAAATGTTTGGCATCAGTACTGACACGTAAAGTTTCCCAATCCT 125  
 Db 4759 GTAAATTTGTTTGAATGTTTGGTATTATTGATACGTAAAGTTTAAATTT 4818

QY 126 CAACCTGCTCCAGCTGATGAGGGGAGGAAGGAGGATTAACCTAGGGGTATGGCGAC 185  
 Db 4819 TAATTTTGTGTTTGTAGTTGATGAGGGGAGGAAGGAGGATTAATTTAGGGGTATGGCGAT 4878  
 QY 186 CAATCTCTGAGTCCCACTGACCACTGACCGCCATCCCGAGCTTGTGCTCACTACCCCA 245  
 Db 4879 TAATTTGAGTTTAAATGATTACGTTTATTTTGTGTTTGTGTTTATTTTATTTT 4938  
 QY 246 ACTCTCCAGAGGAGCAGGATTTTAAAGGGAGCAGGATGCGAGCAACAAACAGACCGCT 305  
 Db 4939 ATTATTTAGAGGAGTAGTTTATTTAAGGGAGTAGGATGAGATAAATAAGACGGTTT 4998  
 QY 306 GGGGATACAACTCTGAGTCCCTCTGAG 332  
 Db 4999 GGGGATATAATTTTGGAGTTTGTGAG 5025

RESULT 12  
 ABV09174/c  
 ID ABV09174 standard; cDNA; 204 BP.  
 XX  
 AC ABV09174;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 XX Human prostate expression marker cDNA 9165.  
 XX  
 DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 XX 23-AUG-2001.  
 XX  
 XX 20-FEB-2001; 2001WO-US05171.  
 XX  
 XX 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 XX Schlegel R, Endege WO, Monahan JE;  
 XX WPI; 2001-662795/76.  
 XX  
 XX Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX  
 PS Claim 1; Page 1442; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 204 BP; 26 A; 62 C; 55 G; 61 T; 0 other;  
 SQ  
 Query Match 35.8%; Score 118.8; DB 23; Length 204;  
 Best Local Similarity 83.3%; Pred No 1.2e-28;  
 Matches 135; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
 QY 171 AGGGTATGGGAGCAACCAATCCTGAGTCCACCACTGACACGCCCATCCCGAGCTTGTG 230  
 Db 191 AGGAAAAGAGCAACAGATCCAGGGAGCATTCACCTGCCCTGTCTCCAAACAGCCTTGTG 132  
 QY 231 CCTCACCCTACCCCACTCCAGAGGAGCAGCTATTATTAAGGGGAGCAGGAGTGCGAA 290  
 Db 131 CCTCACCCTACCCCACTCCAGAGGAGCAGCTATTATTAAGGGGAGCAGGAGTGCGAA 72  
 QY 291 CAAACAGAGCGGCTGGGATACAACTCTGAGTCTCTCTGAG 332  
 Db 71 CAAACAGAGCGGCTGGGATACAACTCTGAGTCTCTCTGAG 30

RESULT 13  
 ABV39325  
 ID ABV39325 standard; cDNA; 481 BP.  
 XX  
 AC ABV39325;  
 XX  
 DT 16-SEP-2002 (first entry)  
 XX  
 XX Human prostate expression marker cDNA 39316.  
 XX  
 DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 XX 23-AUG-2001.  
 XX  
 XX 20-FEB-2001; 2001WO-US05171.  
 XX  
 XX 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 XX Schlegel R, Endege WO, Monahan JE;  
 XX WPI; 2001-662795/76.  
 XX  
 XX Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX  
 PS Claim 1; Page 7980; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

CC patient;  
CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.  
CC Sequence 481 BP; 120 A; 126 C; 142 G; 93 T; 0 other;  
Query Match 35.8%; Score 118.8; DB 23; Length 481;  
Best Local Similarity 83.3%; Pred. NO. 1.7e-28;  
Matches 135; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
YY 171 AGGGGTATGGCGGCAATCTGTGATCCACCACTGACACGCGCCATCCCGCCTTGTG 230  
YY 60 AGGAAAAGAGACAGATCCAGGAGATTTCCTGCTTCCAAACAGCCTTGTG 119  
YY 231 CCTACCTACCCCAACCTCCAGAGGAGCAGCTATTAAAGGGAGCAGGATGCAGAA 290  
YY 120 CCTACCTACCCCAACCTCCAGAGGAGCAGCTATTAAAGGGAGCAGGATGCAGAA 179  
YY 291 CAACAACAGCGGCTGGGATACAACTCTGAGTCCTCTGAG 332  
YY 180 CAACAACAGCGGCTGGGATACAACTCTGAGTCCTCTGAG 221  
RESULT 14  
ABK63588  
ID ABK63588 standard; cDNA; 3330 BP.  
AC ABK63588;  
AT 18-JUN-2002 (first entry)  
DE Rat sequence differentially expressed in response to a hepatotoxin #1495.  
KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
KW differential expression; centrilobular necrosis; steatosis.  
OS Rattus norvegicus.  
XX WO200210453-A2.  
XX 07-FEB-2002.  
XX 30-JUL-2001; 2001WO-US23872.  
XX 31-JUL-2000; 2000US-222040P.  
XX 02-NOV-2000; 2000US-244880P.  
XX 11-MAY-2001; 2001US-290029P.  
XX 15-MAY-2001; 2001US-290645P.  
XX 22-MAY-2001; 2001US-292336P.  
XX 06-JUN-2001; 2001US-295798P.  
XX 13-JUN-2001; 2001US-297457P.  
XX 19-JUN-2001; 2001US-298884P.  
XX 09-JUL-2001; 2001US-303459P.  
XX (GENE-) GENE LOGIC INC.  
XX Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;  
XX WPI; 2002-241625/29.  
XX Predicting toxic effects of compounds or the progression of these toxic  
XX effects by determining the changes in gene expression in tissues or  
XX cells exposed to the toxin and comparing these to gene expression in  
XX unexposed tissues or cells -  
XX Claim 1; Seq ID No 1495; 239pp; English.  
XX The invention relates to methods for predicting toxic effects of  
XX compounds or the progression of these toxic effects by determining the  
XX global changes in gene expression in tissues or cells exposed to the  
XX toxin and comparing these to gene expression in unexposed tissues or  
XX cells. Also included are methods of predicting a toxic effect, preferably the  
XX effect of a compound or progression of a toxic effect, preferably the  
XX hepatotoxicity of a compound, comprising detecting the level of

CC expression in a tissue or cell sample exposed to the compound of two or  
CC more genes listed in the specification, where differential expression of  
CC the genes is indicative of at least one toxic effect or progression.  
CC The method can also be used to identify an agent which modulates the  
CC toxic response and predict cellular pathways that a compound modulates  
CC in a cell. The methods utilize a set of at least two probes (on a solid  
CC support in kit form), where each of the probes comprises a sequence that  
CC specifically hybridizes to a gene listed in the specification, a computer  
CC system comprising a database containing information identifying the  
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
CC set of genes comprising at least two genes listed in the specification,  
CC and a user interface to view the information used to present information,  
CC identifying the expression level in a tissue or cell of at least one gene  
CC listed in the specification. The method is useful for elucidating global  
CC changes in gene expression and for identifying toxicity markers in  
CC tissues or cell exposed to a known toxin. The genes may be used as  
CC toxicity markers in drug screening and toxicity assays. The genes and  
CC gene expression information may be used as diagnostic markers for the  
CC prediction or identification of the physiological state of tissue or cell  
CC sample that has been exposed to a compound or agent. Hepatotoxicity  
CC is characterised by centrilobular necrosis and steatosis. The present  
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
CC which is differentially expressed in response to a hepatotoxic agent.  
XX  
SQ Sequence 3330 BP; 858 A; 860 C; 858 G; 753 T; 1 other;  
Query Match 25.4%; Score 84.4; DB 24; Length 3330;  
Best Local Similarity 64.9%; Pred. NO. 5.8e-17;  
Matches 157; Conservative 0; Mismatches 81; Indels 4; Gaps 2;  
QY 67 CAACAACGCTCGAAATGTGTTGGCATCATGCTACTGACACGTAAGTTTCCCAATCTTC 126  
DB 260 CGAATCAGCTAAGTTTATGATGGCCACACCCATGTTAGGGCTTTTCGGGCCCTC 319  
QY 127 AACTCTGTCTCCAGCTGATGAGGGAGGAAAGGATACCTAGGGGTATGGG--CGA 184  
DB 320 AAGGCTGTCTCCAGCTGTTCGGGGGAAAGGGGAAATTACCCAGGGCTTGGGTATGC 379  
QY 185 CCAATCTGAGTCCACCACTGACACGCGCCA--TCCCGAGCCTTGTGCTACCTACCC 242  
DB 380 CGTCTGTGATCCATTATTGTCACACCCACCTCCCATCTCTGGCTCTCCGATCC 439  
QY 243 CCAACCTCCCGAGGGAGCGCTATTTTAAGGGAGCGAGGTGTCAGAACAAACAGACGG 302  
DB 440 CCAGCCTTCGACAGGGAAGAGCTATTTAAGAGCATTTGGGAGTACAGGAAACAGGCAG 499  
QY 303 CC 304  
DB 500 GC 501  
RESULT 15  
AAF15635  
ID AAF15635 standard; cDNA; 1076 BP.  
AC AAF15635;  
AT 13-MAR-2001 (first entry)  
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:70.  
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
XX Human; prostate cancer; cytosolic; cardiotoxic; immunomodulatory; muscular;  
XX neuroprotective; cytotoxic; cardiotoxic; immunomodulatory; muscular;  
XX vulnary; gastrointestinal; nephrotropic; antiinfective; synaecological;  
XX antibacterial; gene therapy; neural; immune; reproductive; renal;  
XX gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
XX wound; infectious disease; ss.  
XX Homo sapiens.  
OS WO200055174-A1.  
XX 21-SEP-2000.  
PD

Search completed: February 19, 2003, 21:23:33  
Job time : 216.781 secs



Result No.	Score	Query Match	Length	DB	ID	Description
C 1	30.6	9.2	305	4	US-09-328-111-618	Sequence 618, Appl
C 2	30.4	9.2	1549	2	US-08-858-444-1	Sequence 1, Appli
C 3	29.6	8.9	9299	3	US-08-458-434A-7	Sequence 7, Appli
C 4	29.2	8.8	3111	2	US-09-014-969-12	Sequence 12, Appl
C 5	29	8.7	204	4	US-09-506-739-37	Sequence 37, Appl
C 6	29	8.7	1883	1	US-08-202-056-2	Sequence 2, Appli
C 7	29	8.7	1933	1	US-08-076-093A-1	Sequence 1, Appli
C 8	29	8.7	1933	1	US-08-410-451-1	Sequence 1, Appli
C 9	29	8.7	1933	1	US-08-410-455-1	Sequence 1, Appli
C 10	29	8.7	1933	1	US-08-418-919-1	Sequence 1, Appli
C 11	29	8.7	1933	1	US-08-410-453A-2	Sequence 2, Appli
C 12	29	8.7	1933	1	US-08-701-265-1	Sequence 1, Appli
C 13	29	8.7	1933	1	US-08-410-454A-2	Sequence 2, Appli
C 14	29	8.7	1933	2	US-08-284-596-1	Sequence 2, Appli
C 15	29	8.7	1933	2	US-08-410-456A-2	Sequence 2, Appli
C 16	29	8.7	1933	2	US-08-805-478-1	Sequence 1, Appli
C 17	29	8.7	1933	2	US-08-802-627A-1	Sequence 1, Appli
C 18	29	8.7	1933	2	US-08-801-238-1	Sequence 1, Appli
C 19	29	8.7	1933	2	US-08-801-228-1	Sequence 1, Appli
C 20	29	8.7	1933	3	US-09-104-296-1	Sequence 1, Appli
C 21	29	8.7	1933	5	PC1-US94-06380-1	Sequence 1, Appli
C 22	28.6	8.6	3728	1	US-08-111-939-1	Sequence 1, Appli
C 23	28.4	8.6	2992	4	US-09-362-132A-3	Sequence 3, Appli
C 24	28.2	8.5	33	1	US-08-186-895-4	Sequence 4, Appli
C 25	28.2	8.5	722	4	US-08-861-774E-49	Sequence 49, Appl
C 26	28.2	8.5	4258	3	US-07-765-830A-5	Sequence 5, Appli
C 27	28	8.4	997	4	US-09-057-860A-3	Sequence 3, Appli

APPLICANT: Lecka-Czernik, Beata  
TITLE OF INVENTION: No. 5959081el Zinc Binding LIM Protein S2-6  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Benjamin Aaron Adler, Ph.D. J.D.  
STREET: 8011 Candle Lane  
CITY: Houston  
STATE: Texas  
ZIP: 77071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 Mb floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word for Macintosh  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,444  
FILING DATE: May 14, 1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benjamin Aaron Adler, Ph.D.  
REGISTRATION NUMBER: 35,423  
REFERENCE/DOCKET NUMBER: D5988  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (713) 777-2321  
TELEFAX: (713) 777-6908  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1549 bp  
TYPE: nucleic acid  
STRANDEDNESS: single-stranded  
TOPOLOGY: linear  
MOLECULE TYPE:  
DESCRIPTION: c-DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
IMMEDIATE SOURCE:  
POSITION IN GENOME:  
FEATURE:  
PUBLICATION INFORMATION:  
US-08-856-444-1  
Query Match 9.2%; Score 30.4; DB 2; Length 1549;  
Best Local Similarity 57.3%; Pred. No. 1.1;  
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
Qy 111 AGGTTCCCAATCCTCACTCTGTCTCTCCAGCTGATGAGGGGAAGGAAGGATTACCT 170  
Db 530 AAGTGGCCCTCCCGGAGGGTGTCTTCCCAAGAGAGAGGGGAACACACAGGAAGCC 589  
Qy 171 AGGGGTATGGCGACCAATCCTGAGTCCACCACTG 205  
Db 590 AGAGGGGCACAGACCACTGCTGCTACCAACCAACGG 625  
RESULT 3  
US-08-458-434A-7/c  
Sequence 7, Application US/08458434A  
Patent No. 6083690  
GENERAL INFORMATION:  
APPLICANT: Harris Ph.D., Stephen E.  
APPLICANT: Mundy M.D., Gregory R.  
APPLICANT: Gosh-Choudhury Ph.D., Nandini  
APPLICANT: Feng Ph.D., Jian Q.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING  
TITLE OF INVENTION: OSTEOGENIC AGENTS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: James C. Weseman, Esq.

STREET: 401 B. Street, Suite 1700  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,434A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Weseman, James C.  
REGISTRATION NUMBER: 30,507  
REFERENCE/DOCKET NUMBER: P00060USO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 699-3604  
TELEFAX: 619-236-1048  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9299 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-458-434A-7  
Query Match 8.9%; Score 29.6; DB 3; Length 9299;  
Best Local Similarity 59.5%; Pred. No. 5;  
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
Qy 211 CGCCCATCCCGAGCCTTGCTCCTACCTCCACCCCACTCCAGAGGGAGCAGCTATTTA 270  
Db 6318 CCCCCACCCCGCCCTTCCTCCGCCCTCCAGCCCAATTCCACAACTTCCAGCTGGTTA 6259  
Qy 271 AGGGGAGCAGGAGTGCAGAACAA 294  
Db 6258 AGAACAGAGGAGGGGAGAACAGA 6235  
RESULT 4  
US-09-014-969-12  
Sequence 12, Application US/09014969  
Patent No. 5965337  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallie, Edward R.  
APPLICANT: Racie, Lisa A.  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/014,969

FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Spurger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3111 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-014-969-12

Query Match 8.8%; Score 29.2; DB 2; Length 3111;  
Best Local Similarity 57.8%; Pred. No. 3.9;  
Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Y 133 GTCTGCCAGCTGATGAGGGAAGAAAGGATTACCTAGGGGTATGGCGACCAATCCT 192  
b 957 GTGTCGCCGGTGATGACAGCTTGAGATAGAAAGACTACAGGCTGAGCTGCCAATCCC 1016

Y 193 GAGTCACCACTGACCGCCGCTCCCA 222  
b 1017 CATGCCGGGATCTTCCACACCCGCTCTCA 1046

RESULT 5  
US-09-506-729-37/c  
; Sequence 37, Application US/09506729  
; Patent No. 6365352  
; GENERAL INFORMATION:  
; APPLICANT: Yerramilli, Subrahmanyam V.  
; APPLICANT: Prashar, Yatindra  
; APPLICANT: Newberger, Peter  
; APPLICANT: Goguen, Jon  
; APPLICANT: Weisman, Sherman M.  
; TITLE OF INVENTION: A PROCESS TO STUDY CHANGES IN GENE EXPRESSION IN  
; FILE OF INVENTION: GRANULOCYTIC CELLS  
; FILE REFERENCE: 44921-5016-US  
; CURRENT APPLICATION NUMBER: US/09/506.729  
; CURRENT FILING DATE: 2000-02-18  
; EARLIER APPLICATION NUMBER: PCT/US98/17284  
; EARLIER FILING DATE: 1998-08-21  
; EARLIER APPLICATION NUMBER: 60/056,844  
; EARLIER FILING DATE: 1997-08-22  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 37  
; LENGTH: 204  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-506-729-37

Query Match 8.7%; Score 29; DB 4; Length 204;  
Best Local Similarity 57.0%; Pred. No. 1.1;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Y 65 GGCAGAACTGCTGAAATGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAATCC 124  
b 194 GGAACATCTGCTGCCCAATGAGTGGTGGCTGCACATGGCTTTCTAGGGATGCTGATC 135

Y 125 TCAACTCTGCTGCCAGCTGATGAGGGAAGG 157  
b 134 TGCACCCAGCTGGAGCTGCAGAGGGGAAGG 102

RESULT 6  
US-08-202-056-2/c  
; Sequence 2, Application US/08202056

Patent No. 5440021  
GENERAL INFORMATION:  
APPLICANT: Chuntharapai, Anan  
APPLICANT: Hebert, Caroline  
APPLICANT: Kim, Kyung Jin  
APPLICANT: Lee, James  
TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,056  
FILING DATE: 25-FEB-1994  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/677211  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B.  
REGISTRATION NUMBER: 34,659  
REFERENCE/DOCKET NUMBER: 706P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5530  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1883 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-202-056-2

Query Match 8.7%; Score 29; DB 1; Length 1883;  
Best Local Similarity 57.0%; Pred. No. 3.6;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Y 65 GGCAGAACTGCTGAAATGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAATCC 124  
b 1865 GGAACATCTGCTGCCCAATGAGTGGTGGCTGCACATGGCTTTCTAGGGATGCTGATC 1806

Y 125 TCAACTCTGCTGCCAGCTGATGAGGGAAGG 157  
b 1805 TGCAGCCAGCTGGAAGCTGCAGAGGGGAAGG 1773

RESULT 7  
US-08-076-093A-1/c  
; Sequence 1, Application US/08076093A  
; Patent No. 5543503  
; GENERAL INFORMATION:  
; APPLICANT: Chuntharapai, Anan  
; APPLICANT: Lee, James  
; APPLICANT: Hebert, Caroline  
; APPLICANT: Jin Kim, K.  
; TITLE OF INVENTION: Antibodies to Human PF4A Receptors  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA

```

;
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,093A
; FILING DATE: 11-Jun-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P2
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1933 nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-076-093A-1

Query Match 8.7%; Score 29; DB 1; Length 1933;
Best Local Similarity 57.0%; Pred. No. 3.6;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 65 GGCACAACTGCTCAATGTTTGGCATCAGCTACTGACACGTAAAGGTTTCCCAATCC 124
Db 1915 GGAACATCTGCTCCCAATGAGTGGTGTGCTGCATATGCTTTCTAGGGATGCTGATCC 1856

QY 125 TCAACTCTGCTCCAGCTGATGAGGGGAAG 157
Db 1855 TGCACGCCAGCTGGAAGCTGCAGAGGGGAAG 1823

RESULT 8
US-08-410-451-1/c
; Sequence 1, Application US/08410451
; Patent No. 5552284
; GENERAL INFORMATION:
; APPLICANT: Lee, James,
; APPLICANT: Holmes, William E.,
; APPLICANT: Woods, William I.,
; TITLE OF INVENTION: Human PF4A Receptors and Their Use
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,451
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/234494
; FILING DATE: 28-APR-1994
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706C1DS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; APPLICATION NUMBER: 08/234494

```

```

;
; FILING DATE: 28-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706C1D4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1933 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-410-451-1

Query Match 8.7%; Score 29; DB 1; Length 1933;
Best Local Similarity 57.0%; Pred. No. 3.6;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 65 GGCACAACTGCTCAATGTTTGGCATCAGCTACTGACACGTAAAGGTTTCCCAATCC 124
Db 1915 GGAACATCTGCTCCCAATGAGTGGTGTGCTGCATATGCTTTCTAGGGATGCTGATCC 1856

QY 125 TCAACTCTGCTCCAGCTGATGAGGGGAAG 157
Db 1855 TGCACGCCAGCTGGAAGCTGCAGAGGGGAAG 1823

RESULT 9
US-08-410-455-1/c
; Sequence 1, Application US/08410455
; Patent No. 5571702
; GENERAL INFORMATION:
; APPLICANT: Lee, James,
; APPLICANT: Holmes, William E.,
; APPLICANT: Woods, William I.,
; TITLE OF INVENTION: Human PF4A Receptors and Their Use
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,455
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/234494
; FILING DATE: 28-APR-1994
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706C1DS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; APPLICATION NUMBER: 08/234494

```

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
JS-08-410-455-1

Query Match 8.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.6;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
Y 65 GGCAAACTGCTGAAATGCTTTTGGCATCAGCTACTGACACGTAAAGTTTCCCAATCC 124  
b 1915 GGAACATCTGCTGCCCAATGGATGGTGGCTGCACATGCTTTCTAGGATGCTGATGC 1856  
Y 125 TCAACTCTGCTGCTGCACTGATGAGGGGAAGG 157  
b 1855 TGCAGCCAGCTGGAAGCTGCAGAGGGGAAGG 1823

RESULT 10  
US-08-418-919-1/c  
Sequence 1, Application US/08418919  
Patent No. 5633141  
GENERAL INFORMATION:  
APPLICANT: Lee, James,  
APPLICANT: Holmes, William E.,  
APPLICANT: Woods, William I.,  
TITLE OF INVENTION: Human PFA4 Receptors and Their Use  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/418,919  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/234,494  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Hensley, Max D. 27,043  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER: 706  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/266-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-418-919-1

Query Match 8.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.6;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
Y 65 GGCAAACTGCTGAAATGCTTTTGGCATCAGCTACTGACACGTAAAGTTTCCCAATCC 124

Db 1915 GGAACATCTGCTGCCCAATGGATGGTGGCTGCACATGCTTTCTAGGATGCTGATGC 1856  
QY 125 TCAACTCTGCTGCTGCACTGATGAGGGGAAGG 157  
Db 1855 TGCAGCCAGCTGGAAGCTGCAGAGGGGAAGG 1823

RESULT 11  
US-08-410-453A-2/c  
Sequence 2, Application US/08410453A  
Patent No. 5767063  
GENERAL INFORMATION:  
APPLICANT: Lee, James,  
APPLICANT: Holmes, William E.,  
APPLICANT: Woods, William I.,  
TITLE OF INVENTION: Human PFA4 Receptors and Their Use  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/410,453A  
FILING DATE: 24-Mar-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/234494  
FILING DATE: 28-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/677211  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B.  
REGISTRATION NUMBER: 34,659  
REFERENCE/DOCKET NUMBER: P0706C1D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-5530  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-410-453A-2

Query Match 8.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.6;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
QY 65 GGCAAACTGCTGAAATGCTTTTGGCATCAGCTACTGACACGTAAAGTTTCCCAATCC 124  
Db 1915 GGAACATCTGCTGCCCAATGGATGGTGGCTGCACATGCTTTCTAGGATGCTGATGC 1856  
QY 125 TCAACTCTGCTGCTGCACTGATGAGGGGAAGG 157  
Db 1855 TGCAGCCAGCTGGAAGCTGCAGAGGGGAAGG 1823

RESULT 12  
US-08-701-265-1/c  
Sequence 1, Application US/08701265  
Patent No. 5776457  
GENERAL INFORMATION:

APPLICANT: Chuntharapai, Anan  
APPLICANT: Lee, James  
APPLICANT: Hebert, Caroline  
APPLICANT: Jin Kim, K.  
TITLE OF INVENTION: Antibodies to Human PF4A Receptors  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,265  
FILING DATE: 22-AUG-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/076093  
FILING DATE: 11-Jun-1993  
APPLICATION NUMBER: 07/810782  
FILING DATE: 19-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/677211  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B.  
REGISTRATION NUMBER: 34,659  
REFERENCE/DOCKET NUMBER: 706P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5530  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 nucleotides  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-701-265-1

Query Match 8.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.6;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 65 GGCAAACTGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAATCC 124  
DB 1915 GGAACATCTGCTGCCAATGGACTGGTGGCTGCACATGGCTTCTAGGGATGCTGATGC 1856  
QY 125 TCAACTGTCTGCCAGCTGATGAGGGGAAG 157  
DB 1855 TGCAGCCAGCTGGAAGCTGCAGAGGGGAAG 1823

## RESULT 13

US-08-410-454A-2/C  
Sequence 2, Application US/08410454A  
Patent No. 5783415  
GENERAL INFORMATION:  
APPLICANT: Lee, James,  
APPLICANT: Holmes, William B.,  
APPLICANT: Woods, William I.  
TITLE OF INVENTION: Human PF4A Receptors and Their Use  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco

STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/410,454A  
FILING DATE: 24-Mar-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/234494  
FILING DATE: 28-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/677211  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B.  
REGISTRATION NUMBER: 34,659  
REFERENCE/DOCKET NUMBER: P0706C1D3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5530  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-410-454A-2

Query Match 8.7%; Score 29; DB 1; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.6;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 65 GGCAAACTGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAATCC 124  
DB 1915 GGAACATCTGCTGCCAATGGACTGGTGGCTGCACATGGCTTCTAGGGATGCTGATGC 1856  
QY 125 TCAACTGTCTGCCAGCTGATGAGGGGAAG 157  
DB 1855 TGCAGCCAGCTGGAAGCTGCAGAGGGGAAG 1823

## RESULT 14

US-08-284-586-1/C  
Sequence 1, Application US/08284586  
Patent No. 5840856  
GENERAL INFORMATION:  
APPLICANT: Chuntharapai, Anan  
APPLICANT: Lee, James  
APPLICANT: Hebert, Caroline  
APPLICANT: Jin Kim, K.  
TITLE OF INVENTION: Antibodies to Human PF4A Receptors  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/284,586  
FILING DATE:

CLASSIFICATION: 424  
PRIOR APPLICATION DATA: US/08/076,093A  
APPLICATION NUMBER: 11-Jun-1993  
FILING DATE: 07/810782  
FILING DATE: 19-DEC-1991  
PRIOR APPLICATION DATA: 07/677211  
APPLICATION NUMBER: 29-MAR-1991  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B  
REGISTRATION NUMBER: 34,659  
REFERENCE/DOCKET NUMBER: 706P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5530  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 nucleotides  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
IS-08-284-586-1

Query Match 8.7%; Score 29; DB 2; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.6;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

yy 65 GGCAAACTGCTGAAATGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAATCC 124  
yb 1915 GGAACATCTGCTGCCAATGGACTGGTGGTGACATGGCTTTCTAGGATGCTGATGC 1856

yy 125 TCAACTGTCTGCTGCCAGCTGATGAGGGGAAGG 157  
yb 1855 TGCACGCCAGCCTGGAGCTGCAGAGGGGAAGG 1823

RESULT 15  
IS-08-410-456A-2/c  
Sequence 2, Application US/08410456A  
Patent No. 5856457  
GENERAL INFORMATION:  
APPLICANT: Lee, James,  
APPLICANT: Holmes, William E.,  
APPLICANT: Woods, William I.  
TITLE OF INVENTION: Human P4A Receptors and Their Use  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/410,456A  
FILING DATE: 24-Mar-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/234494  
FILING DATE: 28-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/677211  
FILING DATE: 29-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Love, Richard B.  
REGISTRATION NUMBER: 34,659

REFERENCE/DOCKET NUMBER: P0706C1D2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-5530  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1933 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-410-456A-2

Query Match 8.7%; Score 29; DB 2; Length 1933;  
Best Local Similarity 57.0%; Pred. No. 3.6;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 65 GGCAAACTGCTGAAATGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAATCC 124  
Db 1915 GGAACATCTGCTGCCAATGGACTGGTGGTGACATGGCTTTCTAGGATGCTGATGC 1856

Qy 125 TCAACTGTCTGCTGCCAGCTGATGAGGGGAAGG 157  
Db 1855 TGCACGCCAGCCTGGAGCTGCAGAGGGGAAGG 1823

Search completed: February 20, 2003, 01:16:34  
Job time : 49.4287 secs

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic - nucleic search, using sw model

run on: February 19, 2003, 23:17:14 ; Search time 251.835 Seconds  
(without alignments)  
671.453 Million cell updates/sec

title: US-09-808-388-6

perfect score: 332

sequence: 1 Gtaccatttcgacaaacta.....caactctggagtcctctgag 332

scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

searched: 424239 seqs, 254661826 residues

total number of hits satisfying chosen parameters: 848478

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database : Published Applications NA:

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	332	100.0	332	10	US-09-808-388-6
2	271	81.6	271	10	US-09-808-388-5
3	84.4	25.4	3330	10	US-09-917-800A-1495
4	80.8	24.3	1076	10	US-09-925-300-70
5	50	15.1	735	9	US-09-981-353-17
6	41	12.3	41	10	US-09-808-388-3
7	33.8	10.2	2136	10	US-09-862-658-3
8	33.8	10.2	3320	10	US-09-862-658-1
9	31.4	9.5	771	9	US-10-010-920-4
10	31.4	9.5	771	9	US-10-010-920-5
11	31.4	9.5	955	9	US-10-010-920-3
12	30.6	9.2	305	10	US-09-879-536-618
13	30.6	9.2	2885	10	US-09-880-107-3388
14	30.6	9.2	3370	12	US-10-044-090-339
15	30.2	9.1	187	10	US-09-783-590-3051
16	30.2	9.1	340	10	US-09-833-381-1715
17	30.2	9.1	444	10	US-09-864-761-4561
18	30.2	9.1	30676	10	US-09-927-091-8
19	30.2	9.1	45845	10	US-09-927-091-6

#### ALIGNMENTS

##### RESULT 1

US-09-808-388-6

; Sequence 6, Application US/09808388

; Patent No. US20020081719A1

; GENERAL INFORMATION:

; APPLICANT: Massaad, Charbel

; APPLICANT: Berenbaum, Francis

; APPLICANT: Olivier, Jean-Luc

; APPLICANT: Salvat, Colette

; APPLICANT: Berezat, Gilbert

; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them

; TITLE OF INVENTION: their uses

; FILE REFERENCE: ST00010

; CURRENT APPLICATION NUMBER: US/09/808,388

; CURRENT FILING DATE: 2001-09-20

; PRIOR FILING DATE: FR/00/03262

; PRIOR FILING DATE: 2000-03-14

; PRIOR APPLICATION NUMBER: US 60/196,959

; PRIOR FILING DATE: 2000-04-13

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 6

; LENGTH: 332

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE: PPRE/PLA2s hybrid promoter

; OTHER INFORMATION: PPRE/PLA2s hybrid promoter

US-09-808-388-6

Query Match 100.0%; Score 332; DB 10; Length 332;  
Best Local Similarity 100.0%; Pred. No. 3.8e-103;  
Matches 332; Conservative 0; Mismatches 0; Gaps 0;

QY 1 GTACCAATTCGACAAACTAGGTCAAAAGGTCAATCAAACTAGGTCAAAAGGTCAAAATTCGA 60

Db 1 GTACCAATTCGACAAACTAGGTCAAAAGGTCAATCAAACTAGGTCAAAAGGTCAAAATTCGA 60

QY 61 ACGCGGCAAACTGCTGAAATGTTTGGCATCACTACTGACACGTAAAGTTTCCCA 120

Db 61 ACGCGGCAAACTGCTGAAATGTTTGGCATCACTACTGACACGTAAAGTTTCCCA 120

QY 121 ATCTCAACTCTGCTGCCAGCTGATGAGGGAGGAAAGGATTACCTAGGGGTATGG 180

Sequence 142, App

Sequence 142, App

Sequence 4, Appli

Sequence 761, App

Sequence 1, Appli

Sequence 42, Appli

Sequence 116, App

Sequence 474, App

Sequence 1, Appli

Sequence 9949, Ap

Sequence 120, App

Sequence 3854, Ap

Sequence 2747, Ap

Sequence 277, App

Sequence 4, Appli

Sequence 5, Appli

Sequence 1, Appli

Sequence 2124, Ap

Sequence 829, App

Sequence 1083, Ap

Sequence 829, App

Sequence 1083, Ap

Sequence 829, App

Sequence 1083, Ap



Db 121 ATCTCACTCTGTCTGCCAGTGTAGGGGAGGAAGGATTACCTAGGGGTATGG 180  
QY 181 GCACCAATCTGTAGTCCACCACTGACACGCCCATCCAGCGCTTGTGCTCACCCTAC 240  
Db 181 GCACCAATCTGTAGTCCACCACTGACACGCCCATCCAGCGCTTGTGCTCACCCTAC 240  
QY 241 CCCCACTCTCCAGAGGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAAGAC 300  
Db 241 CCCCACTCTCCAGAGGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAAGAC 300  
QY 301 GGCTGGGATACAACTCTGGAGTCTCTGAG 332  
Db 301 GGCTGGGATACAACTCTGGAGTCTCTGAG 332  
RESULT 2  
US-09-808-388-5  
; Sequence 5, Application US/09808388  
; Patent No. US20020081719A1  
; GENERAL INFORMATION:  
; APPLICANT: Massaad, Charbel  
; APPLICANT: Berenbaum, Francis  
; APPLICANT: Olivier, Jean-Luc  
; APPLICANT: Salvat, Colette  
; APPLICANT: Perezziat, Gilbert  
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
; FILE REFERENCE: ST00010  
; CURRENT APPLICATION NUMBER: US/09/808,388  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: US 60/196,959  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 271  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fragment of the PLA2s promoter  
US-09-808-388-5  
Query Match 81.8%; Score 271; DB 10; Length 271;  
Best Local Similarity 100.0%; Pred. No. 1.8e-82;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 62 CGCGGCAAACTGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAA 121  
Db 1 CGCGGCAAACTGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAA 60  
QY 122 TCCTCAACTGTCTGCCAGCTGATAGGGGAGGAAGGATACCTAGGGGTATGG 181  
Db 61 TCCTCAACTGTCTGCCAGCTGATAGGGGAGGAAGGATACCTAGGGGTATGG 120  
QY 182 CGACCAATCTGTAGTCCACCACTGACACGCCCATCCAGCGCTTGTGCTCACCCTACC 241  
Db 121 CGACCAATCTGTAGTCCACCACTGACACGCCCATCCAGCGCTTGTGCTCACCCTACC 180  
QY 242 CCCCACTCTCCAGAGGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAAGAC 301  
Db 181 CCCCACTCTCCAGAGGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAAGAC 240  
QY 302 GCCTGGGATACAACTCTGGAGTCTCTGAG 332  
Db 241 GCCTGGGATACAACTCTGGAGTCTCTGAG 271  
RESULT 3  
US-09-917-800A-1495  
; Sequence 1495, Application US/09917800A

; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1495  
; LENGTH: 3330  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 X51529  
US-09-917-800A-1495  
Query Match 25.4%; Score 84.4; DB 10; Length 3330;  
Best Local Similarity 54.9%; Pred. No. 1.7e-18;  
Matches 157; Conservative 0; Mismatches 81; Indels 4; Gaps 2;  
QY 67 CAAAACCTGCTGAAATGTGTTTGGCATCAGCTACTGACACGTAAGGTTTCCCAATCTC 126  
Db 260 CGAAATCAGCTAAAGTTATGATGCGCCACACCCCATGATGAGGGCTTTTCCGGCCCTC 319  
QY 127 AACTCTGTCTGCCAGCTGATAGGGGAGGAAGGATACCTAGGGGTATGG--CGA 184  
Db 320 AAGGCTGTCTGCCAGCTGTGCGGGGAGAAAGGGGAAATACCCAGGGCGTTGGGTATGC 379  
QY 185 CCAATCTGAGTCCACCACTGACACGCCCA--TCCCCAGCCTTGTGCTCAGCTACCC 242  
Db 380 CCCTCTGTGAATCAATATTGCGCCACACCCACCTCCCATCCTGTGGCTCTCCGATCC 439  
QY 243 CCAACTCTCCAGAGGGAGCAGCTATTATTAAGGGAGCAGGAGTGCAGAACAAACAAGAC 302  
Db 440 CCAGCCTCTCAGAGGAGAGCTATTATTAAGAGCATTTGGAGTACAGGAAACAAAGGCG 499  
QY 303 CC 304  
Db 500 GC 501  
RESULT 4  
US-09-925-300-70  
; Sequence 70, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

```
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 70
LENGTH: 1076
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
LOCATION: (911)
OTHER INFORMATION: n equals a,t,g, or c
S-09-925-300-70

Query Match      24.3%; Score 80.8; DB 10; Length 1076;
Best Local Similarity 95.3%; Pred. No. 1.6e-17;
Matches 82; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

y 247 CTCCTCCAGAGGAGCGAGCTATTAAAGGAGGAGGAGTCCAGAAACAACAGACGGCGCTG 306
|||
b 2 CCAACAGAGGAGCGAGCTATTAAAGGAGGAGGAGTCCAGAAACAACAGACGGCGCTG 61
|||
y 307 GGGATACAACTCTGGAGTCTCTGAG 332
|||
b 62 GGGATACAACTCTGGAGTCTCTGAG 87
|||

RESULT 5
S-09-981-353-17
Sequence 17, Application US/09981353
Patent No. US20020160382A1
GENERAL INFORMATION:
APPLICANT: Lasek, Amy W.
APPLICANT: Jones, David A.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0038 US
CURRENT APPLICATION NUMBER: US/09/981,353
CURRENT FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 17
LENGTH: 735
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020160382A1 474322.36
LOCATION: 388
OTHER INFORMATION: a, t, c, g, or other
US-09-981-353-17

Query Match      15.1%; Score 50; DB 9; Length 735;
Best Local Similarity 100.0%; Pred. No. 3.8e-07;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 283 GTGCAGAACAAACAGACGGCGCTGGGATACAACTCTGGAGTCTCTGAG 332
|||
yb 1 GTGCAGAACAAACAGACGGCGCTGGGATACAACTCTGGAGTCTCTGAG 50
|||

RESULT 6
US-09-808-388-3
Sequence 3, Application US/09808388
Patent No. US20020081719A1
GENERAL INFORMATION:
APPLICANT: Massaad, Charbel
APPLICANT: Berenbaum, Francis

FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/808,388
CURRENT FILING DATE: 2001-09-20
PRIOR APPLICATION NUMBER: FR/00/03262
PRIOR FILING DATE: 2000-03-14
PRIOR APPLICATION NUMBER: US 60/196,959
PRIOR FILING DATE: 2000-04-13
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patent In version 3.0
SEQ ID NO 3
LENGTH: 41
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PPRE element
US-09-808-388-3

Query Match      12.3%; Score 41; DB 10; Length 41;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 CAAACTAGGTCAAAGGTCAATCAAACTAGGTCAAAGGTCA 53
|||
Db 1 CAAACTAGGTCAAAGGTCAATCAAACTAGGTCAAAGGTCA 41
|||

RESULT 7
US-09-862-658-3
Sequence 3, Application US/09862658
Patent No. US20020137101A1
GENERAL INFORMATION:
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 46638, A NOVEL HUMAN LIPOXYGENASE FAMILY
TITLE OF INVENTION: MEMBER AND USES THEREOF
FILE REFERENCE: 10448-053001
CURRENT APPLICATION NUMBER: US/09/862,658
CURRENT FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/205,675
PRIOR FILING DATE: 2000-05-19
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 2136
TYPE: DNA
ORGANISM: Homo sapiens
US-09-862-658-3

Query Match      10.2%; Score 33.8; DB 10; Length 2136;
Best Local Similarity 53.4%; Pred. No. 0.21;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 129 CTCTGTCTCTGCCAGCTGATGAGGGAAGGAAGGAGTATCACTAGGGGTATGGGACCAA 188
|||
Db 1752 CTGTGCCAGCAGCTGTCTGTCAACAGTGGCAGCATGACTTGGGCTGGATGCCAA 1811
|||
QY 189 TCCTGAGTCCAACTACCAACAGCCGATCCCAAGCCCTTGCGCTCACCCTACCCCAACC 248
|||
Db 1812 TGCTCCATCATCATGAGGAGCGAGCCGCCCAAGCCCAAGGAGGAGCCACCCTGAAGAC 1871
|||
QY 249 TCCAGAGGGAGC 261
|||
Db 1872 TTACCTAGACACC 1884
|||

RESULT 8
US-09-862-658-1
Sequence 1, Application US/09862658
Patent No. US20020137101A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 46638, A NOVEL HUMAN LIPOXYGENASE FAMILY
; FILE REFERENCE: 10448-053001
; CURRENT APPLICATION NUMBER: US/09/862,658
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/205,675
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3320
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (459)...(2591)
US-09-862-658-1

Query Match          10.2%; Score 33.8; DB 10; Length 3320;
Best Local Similarity 53.4%; Pred. No. 0.26;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 129 CTCTGCTCCGACGCTGATGAGGGGAGGAAGGATTAACCTAGGGGTATGGGCGACCAA 188
DB 2210 CTCTGCTCCGACGCTGCTGTCAACAGTGGGCGACATGACTTTGGGGCTTGGATGCCCAA 2269
QY 189 TCTGTAGTCCACCACTGACACGACGACCCATCCCGCTTGTGCTTCACTACCCGCCAAC 248
DB 2270 TGTCTATCATCATGAGGACGACCCCGCCAGACCAAGGGACCCACCCCTGAAGAC 2329
QY 249 TCCAGAGGAGC 261
DB 2330 TTACCTAGACACC 2342

RESULT 9
US-10-010-920-4/c
; Sequence 4, Application US/10010920
; Publication No. US20030027165A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Alternatively spliced polk nucleotide and amino acid sequences
; FILE REFERENCE: 98,723-E3
; CURRENT APPLICATION NUMBER: US/10/010,920
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,649
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-010-920-4

Query Match          9.5%; Score 31.4; DB 9; Length 771;
Best Local Similarity 57.7%; Pred. No. 0.82;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 236 CCTACCCCAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTCCAGAACAAAC 295
DB 578 CCTACCTCCGGCTCTCCCGGCTGACGAGGAGTAAAGGAGGAGGAGGAGGAGGAGGAG 519
QY 296 AAGACGGCTGGGATACACTCTGGAGTCTCTCTGAG 332
DB 518 AGGCGGGGTAGGATGAGCTGTGCTGCATTTCTGGG 482

RESULT 10
US-10-010-920-5
; Sequence 5, Application US/10010920
; Publication No. US20030027165A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Alternatively spliced polk nucleotide and amino acid sequences
; FILE REFERENCE: 98,723-E3
; CURRENT APPLICATION NUMBER: US/10/010,920
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,649
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-010-920-5

Query Match          9.5%; Score 31.4; DB 9; Length 771;
Best Local Similarity 57.7%; Pred. No. 0.82;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 236 CCTACCCCAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTCCAGAACAAAC 295
DB 194 CCTACCTCCGGCTCTCCCGGCTGACGAGGAGTAAAGGAGGAGGAGGAGGAGGAGGAG 253
QY 296 AAGACGGCTGGGATACACTCTGGAGTCTCTCTGAG 332
DB 254 AGGCGGGGTAGGATGAGCTGTGCTGCATTTCTGGG 290

RESULT 11
US-10-010-920-3/c
; Sequence 3, Application US/10010920
; Publication No. US20030027165A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; TITLE OF INVENTION: Alternatively spliced polk nucleotide and amino acid sequences
; FILE REFERENCE: 98,723-E3
; CURRENT APPLICATION NUMBER: US/10/010,920
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,649
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 955
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-010-920-3

Query Match          9.5%; Score 31.4; DB 9; Length 955;
Best Local Similarity 57.7%; Pred. No. 0.91;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 236 CCTACCCCAACCTCCAGAGGAGCAGCTATTTAAGGGGAGCAGGAGTCCAGAACAAAC 295
DB 727 CCTACCTCCGGCTCTCCCGGCTGACGAGGAGTAAAGGAGGAGGAGGAGGAGGAGGAG 668
QY 296 AAGACGGCTGGGATACACTCTGGAGTCTCTCTGAG 332
DB 667 AGGCGGGGTAGGATGAGCTGTGCTGCATTTCTGGG 631

RESULT 12
US-09-879-536-618/c
; Sequence 618, Application US/09879536
; Patent No. US20030144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.

```

APPLICANT: Astle, Jon H.  
APPLICANT: Burgess, Christopher C.  
APPLICANT: Bushnell, Steven E.  
APPLICANT: Carroll III, Eddie  
APPLICANT: Catino, Theodore J.  
APPLICANT: Derti, Adnan  
APPLICANT: Ford, Donna M.  
APPLICANT: Lewis, Marcia E.  
APPLICANT: Monahan, John E.  
APPLICANT: Schlegel, Robert  
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
FILE REFERENCE: CCD-257 (US)  
CURRENT APPLICATION NUMBER: US/09/879,536  
CURRENT FILING DATE: 2001-09-21  
PRIOR APPLICATION NUMBER: US 60/088,901  
PRIOR FILING DATE: 1998-06-10  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 618  
LENGTH: 305  
TYPE: DNA  
ORGANISM: Homo sapiens  
S-09-879-536-618

Query Match 9.2%; Score 30.6; DB 10; Length 305;  
Best Local Similarity 56.4%; Pred. No. 0.96; 44; Indels 0; Gaps 0;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Y 64 CGGCAAACTGCCTGAATGTGTTTGGCATCAGTACTGACACGTAAGTTTCCCAATC 123  
b 217 CGCCAAAATAACACGATGTTGTAAACATCCCCCAGTGGGGCTAGAAATCCCCCATG 158

Y 124 CTCAACTCTGCTCCGACGCTGATGAGGGAGGAAAGGGA 164  
b 157 GTGACCTGTGACCTGCTCCCTGAGACAGGGGAGGCCAGGCA 117

RESULT 13  
S-09-880-107-3388/c  
Sequence 3388, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3388  
LENGTH: 2885  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 U53347  
IS-09-880-107-3388

Query Match 9.2%; Score 30.6; DB 10; Length 2885;  
Best Local Similarity 56.4%; Pred. No. 3;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Y 64 CGGCAAACTGCCTGAATGTGTTTGGCATCAGTACTGACACGTAAGTTTCCCAATC 123  
b 2633 CGCCAAAATAACACGATGTTGTAAACATCCCCCAGTGGGGCTAGAAATCCCCCATG 2574

APPLICANT: 124 CTCAACTCTGCTCCGACGCTGATGAGGGAGGAAAGGGA 164  
Db 2573 GTGACCTGTGACCTGCTCCCTGAGACAGGGAGGCCAGGCA 2533

RESULT 14  
US-10-044-090-339/c  
Sequence 339, Application US/10044090  
Patent No. US20020137081A1  
GENERAL INFORMATION:  
APPLICANT: Olga Bandman  
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
FILE REFERENCE: PA-0028 US  
CURRENT APPLICATION NUMBER: US/10/044,090  
CURRENT FILING DATE: 2002-01-09  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: PERL Program  
SEQ ID NO 339  
LENGTH: 3370  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: Incyte ID No. US20020137081A1 346535.6  
US-10-044-090-339

Query Match 9.2%; Score 30.6; DB 12; Length 3370;  
Best Local Similarity 56.4%; Pred. No. 3.2; 44; Indels 0; Gaps 0;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Y 64 CGGCAAACTGCCTGAATGTGTTTGGCATCAGTACTGACACGTAAGTTTCCCAATC 123  
Db 3141 CGCCAAAATAACACGATGTTGTAAACATCCCCCAGTGGGGCTAGAAATCCCCCATG 3082

Y 124 CTCAACTCTGCTCCGACGCTGATGAGGGAGGAAAGGGA 164  
Db 3081 GTGACCTGTGACCTGCTCCCTGAGACAGGGAGGCCAGGCA 3041

RESULT 15  
US-09-783-590-3051/c  
Sequence 3051, Application US/09783590  
Patent No. US2002010850A1  
GENERAL INFORMATION:  
APPLICANT: Dillon, Patrick J.  
APPLICANT: Haseltine, William A.  
APPLICANT: Li, Haodong  
APPLICANT: Rosen, Craig A.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
FILE REFERENCE: PO-16.2C1  
CURRENT APPLICATION NUMBER: US/09/783,590  
CURRENT FILING DATE: 2000-02-15  
PRIOR APPLICATION NUMBER: 08/420,856  
PRIOR FILING DATE: 1995-04-12  
PRIOR APPLICATION NUMBER: 08/346,731  
PRIOR FILING DATE: 1994-11-21  
NUMBER OF SEQ ID NOS: 12485  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3051  
LENGTH: 187  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (10)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc.feature  
LOCATION: (19)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc.feature  
LOCATION: (157)  
OTHER INFORMATION: n equals a,t,g, or c

```
; NAME/KEY: misc feature
; LOCATION: (159)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (164)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3051

Query Match          9.1%; Score 30.2; DB 10; Length 187;
Best Local Similarity 58.2%; Pred. No. 1;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 231 CCTCACCTACCCCAACCTCCAGAGGGAGCAGCTATTTAAGGGGAGCAGGAGTCAGAA 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 CCACGGCAGCCAGGACACCCGCTAGGGAGGAGCAGGTTACCGAGGGCGGGAGAGCGGCA 74

QY 291 CAAACAGACGGCCTGGGATACAACTCTGG 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73 CAAACGGCAGACCTCGGGACTCCACCCGGG 43
```

Search completed: February 20, 2003, 07:01:56  
Job time : 256.835 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

DN nucleic - nucleic search, using sw model

Run on: February 19, 2003, 19:33:48 ; Search time 1615.42 Seconds  
(without alignments)  
3328.484 Million cell updates/sec

Title: US-09-808-388-6  
Perfect score: 332  
Sequence: 1 gtcccaattgcacaaacta.....caactctgagtcctctgag 332

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estma:\*

5: em\_estov:\*

6: em\_estopl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	33.7	733	12	BF674954 602138032
2	81.2	24.5	980	14	BQ651334 AGENCOURT
3	53	16.0	902	14	BQ650118 AGENCOURT
4	53	16.0	910	14	BQ648623 AGENCOURT
5	53	16.0	935	14	BQ650042 AGENCOURT
6	40.6	12.2	600	12	BG803927 0243-51 M

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

7	11.1	535	12	BG566944
8	11.1	592	12	BG562803
9	11.1	630	12	BG573704
10	11.1	650	12	BG564662
11	11.1	730	12	BG621542
12	11.1	785	12	BG567864
13	11.1	788	12	BF677811
14	11.1	1101	17	CNS006BP
15	11.0	1083	14	BQ58613
16	10.8	590	14	BQ807326
17	10.7	544	10	AV844441
18	10.7	1101	17	CNS00DRA
19	10.6	1087	13	BM458295
20	10.6	1201	17	CNS015WQ
21	10.5	357	14	N75549
22	10.5	709	9	AL525279
23	10.4	990	17	CNS078WN
24	10.4	1026	13	BM044976
25	10.3	907	17	CNS05515
26	10.2	780	12	BF981904
27	10.2	368	9	AA565892
28	10.2	902	17	CNS00462
29	10.1	588	9	AA673175
30	10.1	812	14	BQ229858
31	10.1	1047	13	BQ944238
32	10.1	1059	13	BM470242
33	10.1	1068	13	BM544574
34	10.1	507	12	BF754755
35	10.1	970	17	CNS06UR1
36	10.0	482	17	AZ854111
37	10.0	562	12	BF079717
38	10.0	845	9	AL571180
39	10.0	1096	14	BM913074
40	9.9	266	12	BG182293
41	9.9	272	12	BF888447
42	9.9	343	13	BI643736
43	9.9	696	17	CNS02BW7
44	9.9	897	17	AZ185982
45	9.9	941	13	BM451198

ALIGNMENTS

RESULT 1  
BF674954 602138032F1 NIH\_MGC\_83 Homo sapiens cDNA clone IMAGE:4274550 5', linear EST 21-DEC-2000  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BF674954 602138032F1 NIH\_MGC\_83 Homo sapiens cDNA clone IMAGE:4274550 5', linear EST 21-DEC-2000  
mRNA sequence.  
BF674954.1 GI:11948849  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 733)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: CLONETECH Laboratories, Inc.  
cDNA Library Preparation: CLONETECH Laboratories, Inc.  
cDNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LICM1091 row: p column: 07  
High quality sequence stop: 649.  
Location/Qualifiers  
1. 733

FEATURES  
source



```

RESULT 4
BQ648623
LOCUS
DEFINITION
AGENCOURT 8297960 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6270023
5', mRNA sequence.
ACCESSION
BQ648623
VERSION
BQ648623.1 GI:21772795
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 910)
REFERENCE
NIH-MGC http://mgs.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM2444 row: h column: 24
High quality sequence stop: 605.
Location/Qualifiers
1. .910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6270023"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 237 a 265 c 224 g 184 t
ORIGIN
Query Match 16.0%; Score 53; DB 14; Length 910;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 GGAGTCGAGACAAACAGCGCTGGGATACAACTCTGGAGTCCTCTGAG 332
|||||
Db 1 GGAGTCGAGACAAACAGCGCTGGGATACAACTCTGGAGTCCTCTGAG 53

RESULT 6
BQ803927
LOCUS
DEFINITION
AGENCOURT 8302541 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271292
5', mRNA sequence.
ACCESSION
BQ803927
VERSION
BQ803927.1 GI:17950840
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 600)
REFERENCE
Mu.X., Zhao.S., Pershad,R., Hsieh,T.-F., Scarpa.A., Wang,S.W.,
White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
JOURNAL
Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE
21671825
COMMENT
Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.
Location/Qualifiers
1. .600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pANP10 (Gibco); Cloned unidirectionally."
BASE COUNT 261 a 257 c 226 g 191 t
ORIGIN
Query Match 16.0%; Score 53; DB 14; Length 935;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 GGAGTCGAGACAAACAGCGCTGGGATACAACTCTGGAGTCCTCTGAG 332
|||||
Db 1 GGAGTCGAGACAAACAGCGCTGGGATACAACTCTGGAGTCCTCTGAG 53

RESULT 5
BQ650042
LOCUS
DEFINITION
AGENCOURT 8302541 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271292
5', mRNA sequence.
ACCESSION
BQ650042
VERSION
BQ650042.1 GI:21774214
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 935)
REFERENCE
NIH-MGC http://mgs.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)

```

COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LNCM2447 row: m column: 21  
High quality sequence stop: 691.  
Location/Qualifiers

1. .935  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6271292"  
/clone\_lib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCACGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

FEATURES

source

BASE COUNT

261 a 257 c 226 g 191 t

ORIGIN

Query Match 16.0%; Score 53; DB 14; Length 935;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 GGAGTCGAGACAAACAGCGCTGGGATACAACTCTGGAGTCCTCTGAG 332

|||||

Db 1 GGAGTCGAGACAAACAGCGCTGGGATACAACTCTGGAGTCCTCTGAG 53

RESULT 6

BQ803927

LOCUS

DEFINITION

AGENCOURT 8302541 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6271292

5', mRNA sequence.

ACCESSION

BQ803927

VERSION

BQ803927.1 GI:17950840

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 600)

REFERENCE

Mu.X., Zhao.S., Pershad,R., Hsieh,T.-F., Scarpa.A., Wang,S.W.,

White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.

Gene expression in the developing mouse retina by EST sequencing

and microarray analysis

JOURNAL

Nucleic Acids Res. 29 (24), 4983-4993 (2001)

MEDLINE

21671825

COMMENT

Contact: Klein WH

Department of Biochemistry and Molecular Biology

University of Texas M.D. Anderson Cancer Center

Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA

Tel: 713 792 3646

Fax: 713 790 0329.

Location/Qualifiers

1. .600

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"

/tissue\_type="neural retina"

/dev\_stage="embryonic day 14.5 post-fertilization"

/notes="Vector: pANP10 (Gibco); Cloned unidirectionally."

FEATURES

source



BASE COUNT	ORIGIN
146 a	130 c
150 g	109

ACCESSION  
VERSION

BASE COUNT

ACCESSION EG573704  
VERSION EG573704.1 GI:13581357

159 a 147 c 163 g 123 t

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Eutharia; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutharia; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 630)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1579 row: i column: 14
High quality sequence stop: 625.
FEATURES
Location/Qualifiers
1..630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4722133"
/clone_lib="NIH MGC 79"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: placenta; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccgctggcc); Site 2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGCCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3'
(where B = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
BASE COUNT 174 a 158 c 170 g 128 t
ORIGIN
Query Match 11.1%; Score 37; DB 12; Length 630;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 296 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 332
Db 3 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 39
RESULT 10
BG564662 650 bp mRNA linear EST 10-APR-2001
LOCUS 602584162F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711838 5',
DEFINITION mRNA sequence.
ACCESSION BG564662
VERSION BG564662.1 GI:13572314
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutharia; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 650)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1588 row: i column: 08
High quality sequence stop: 730.
FEATURES
Location/Qualifiers
1..730
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4730959"
/clone_lib="NIH MGC 79"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: placenta; Vector: pDNR-LIB (Clontech);
Site 1: SfiI (ggccgctggcc); Site 2: SfiI (ggccattggcc)
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGCCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCACATG-dt(30)BN-3'
(where B = A, C, G, or T). Average

```

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM1552 row: 1 column: 15

High quality sequence stop: 625.

FEATURES  
source

1..650

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4711838"

/clone\_lib="NIH MGC 76"

/lab\_host="DH10B (TI phage-resistant)"

/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:

SfiI (ggccgctggcc); Site 2: SfiI (ggccattggcc); 5' and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CACGCCATTATGCCC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCGGCGGCACATG-dt(30)BN-3' (where B = A, A, C,

G, or T). Average insert size 1.85

kb (range 1.0-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH MGC Library."

BASE COUNT 169 a 175 c 166 g 140 t

ORIGIN

Query Match 11.1%; Score 37; DB 12; Length 650;

Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 332

Db 2 AAGACGGCTGGGATACAACTCTGGAGTCTCTGAG 38

RESULT 11

LOCUS BG621542

DEFINITION 602617129F1 NIH\_MGC\_79 Homo sapiens cDNA clone IMAGE:4730959 5',

ACCESSION BG621542

VERSION BG621542.1 GI:13672913

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutharia; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 730)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM1588 row: i column: 08

High quality sequence stop: 730.

FEATURES

Location/Qualifiers

1..730

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4730959"

/clone\_lib="NIH MGC 79"

/lab\_host="DH10B (TI phage-resistant)"

/note="Organ: placenta; Vector: pDNR-LIB (Clontech);

Site 1: SfiI (ggccgctggcc); Site 2: SfiI (ggccattggcc)

; 5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGCCC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGGCGGCACATG-dt(30)BN-3'

(where B = A, C, G, or T). Average

insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

BASE COUNT 196 a 195 c 184 g 155 t  
ORIGIN

Query Match 11.1%; Score 37; DB 12; Length 730;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 AAGACGGCTGGGGATACAACTCTGGAGTCTCTGAG 332  
|||||  
Db 3 AAGACGGCTGGGGATACAACTCTGGAGTCTCTGAG 39  
|||||

RESULT 12  
BG567864 785 bp mRNA linear EST 10-APR-2001  
LOCUS 60258685F1 NIH\_MGC\_76 Homo sapiens cDNA clone IMAGE:4715612 5',  
DEFINITION mRNA sequence.

ACCESSION BG567864  
VERSION BG567864.1 GI:13575517  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 785)  
NIH-MGC http://mgi.nci.nih.gov/

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: L1CM1562 row: i column: 21  
High quality sequence stop: 764.

FEATURES  
Location/Qualifiers

1. .785  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4715612"  
/clone\_lib="NIH\_MGC\_76"  
/lab\_host="DH10B (TI phage-resistant)"

/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCGGCGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

BASE COUNT 205 a 225 c 187 g 168 t  
ORIGIN

Query Match 11.1%; Score 37; DB 12; Length 785;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 AAGACGGCTGGGGATACAACTCTGGAGTCTCTGAG 332  
|||||  
Db 3 AAGACGGCTGGGGATACAACTCTGGAGTCTCTGAG 39  
|||||

RESULT 13  
BF677811  
LOCUS 602085322F1 NIH\_MGC\_83 Homo sapiens cDNA clone IMAGE:4249492 5',  
DEFINITION mRNA sequence.

ACCESSION BF677811  
VERSION BF677811.1 GI:11951706  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 788)  
NIH-MGC http://mgi.nci.nih.gov/

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: L1CM1089 row: i column: 05  
High quality sequence stop: 668.

FEATURES  
Location/Qualifiers

1. .788  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4249492"  
/clone\_lib="NIH\_MGC\_83"  
/lab\_host="DH10B (TI phage-resistant)"

/note="Organ: prostate; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCGGCGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 207 a 203 c 203 g 175 t  
ORIGIN

Query Match 11.1%; Score 37; DB 12; Length 788;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 AAGACGGCTGGGGATACAACTCTGGAGTCTCTGAG 332  
|||||  
Db 2 AAGACGGCTGGGGATACAACTCTGGAGTCTCTGAG 38  
|||||

RESULT 14  
CNS006BP 1101 bp DNA linear GSS 03-JUN-1999

LOCUS CNS006BP  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR33018 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL064052  
VERSION AL064052.1 GI:4944232  
KEYWORDS GSS.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)  
Genoscope.  
Direct Submission



score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	944	100.0	944	6	AX251579	AX251579 Sequence
2	237.6	25.2	32492	10	MUSPACOLL	M65161 Mouse pro-a
3	229.2	24.3	183204	2	AC113444	AC113444 Mus muscu
4	216.6	22.9	7175	6	AX114853	AX114853 Sequence
5	172.8	18.3	823	10	RATCOLIIE	I48618 Rattus norv
6	171	18.1	197631	2	AC098511	AC098511 Rattus no
7	170.6	18.1	861	10	MUSCOL2A1A	M63710 Mus musculu
8	138.6	14.7	1404	10	RATCG2A1A	M10613 Rat alpha-1
9	124.6	13.2	6723	9	HUMCOLI1	M60299 Human alpha
10	124	13.1	193561	9	AC004801	AC004801 Homo sapi
11	122.4	13.0	5887	9	HSPROCOE1	X58709 Human type
12	122.4	13.0	31001	9	HUMCOL2A1Z	L10347 Human pro-a
13	114.2	12.1	282	6	AX034871	AX034871 Sequence
14	113.8	12.1	758	9	HUMCOLARAB	M25698 Human alpha
15	113	12.0	244	6	AX034878	AX034878 Sequence
16	112.8	11.9	246	6	AX034879	AX034879 Sequence
17	112.8	11.9	283	6	AX034870	AX034870 Sequence
18	112.6	11.9	241	6	AX034888	AX034888 Sequence
19	111.4	11.8	235	6	AX034880	AX034880 Sequence
20	111.4	11.8	247	6	AX034881	AX034881 Sequence
21	111.4	11.8	256	6	AX034882	AX034882 Sequence
22	111.4	11.8	256	6	AX034883	AX034883 Sequence
23	111.4	11.8	282	6	AX034872	AX034872 Sequence
24	111.4	11.8	282	6	AX034873	AX034873 Sequence
25	111.4	11.8	282	6	AX034874	AX034874 Sequence
26	111.4	11.8	282	6	AX114872	AX114872 Sequence
27	109.2	11.6	8062	6	AX114871	AX114871 Sequence
28	109.2	11.6	8153	6	AX114841	AX114841 Sequence
29	104.4	11.1	120	6	AX114841	AX114841 Sequence
30	104.4	11.1	120	6	AX114841	AX114841 Sequence
31	104.4	11.1	179	9	HUNSRU30S	M34493 Human 30S s
32	104.4	11.1	179	12	SYNLBRNA	M15755 Modified ad
33	104.4	11.1	471	14	ADRRNASEQA	M30268 Adenovirus
34	104.4	11.1	10332	6	A83180	A83180 Sequence 13
35	104.4	11.1	11570	14	AD5001	X02996 Adenovirus
36	104.4	11.1	32798	6	AX382187	AX382187 Sequence
37	104.4	11.1	33007	12	AF323988	AF323988 EGFP expr
38	104.4	11.1	33592	6	AX084504	AX084504 Sequence
39	104.4	11.1	33699	6	AX084506	AX084506 Sequence
40	104.4	11.1	33988	6	AX084517	AX084517 Sequence
41	104.4	11.1	34303	6	AR091536	AR091536 Sequence
42	104.4	11.1	34303	6	AR102229	AR102229 Sequence
43	104.4	11.1	34341	6	AX084505	AX084505 Sequence
44	104.4	11.1	34448	6	AX084507	AX084507 Sequence
45	104.4	11.1	34737	6	AX084518	AX084518 Sequence

ALIGNMENTS

RESULT 1	AX251579	AX251579	Sequence	7	from Patent	WO0168845.	944 bp	DNA	linear	PAT 05-OCT-2001
LOCUS	AX251579	AX251579	Sequence	7	from Patent	WO0168845.				
DEFINITION	AX251579	AX251579	Sequence	7	from Patent	WO0168845.				
ACCESSION	AX251579	AX251579	Sequence	7	from Patent	WO0168845.				
VERSION	AX251579.1	GI:15985002								
KEYWORDS			synthetic construct.							
SOURCE			synthetic construct							
ORGANISM			artificial sequences.							
REFERENCE			1 (bases 1 to 944)							
AUTHORS			Massaad, C., Berenbaum, F., Olivier, J.L., Salvat, C. and Bereziat, C.							
TITLE			Inflammation-inducible hybrid promoters, vectors containing same							
JOURNAL			Patent: WO 0168845-A 7 20-SEP-2001;							

Pred. No. is the number of results predicted by chance to have a

FEATURES  
 source  
 Adventis Pharma S.A. (FR)  
 Location/Qualifiers  
 1..944  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="sequence conferant la specificite d'expression"  
 BASE COUNT 165 a 335 c 232 g 212 t  
 RIGIN  
 Query Match 100.0%; Score 944; DB 6; Length 944;  
 Best Local Similarity 100.0%; Pred. No. 1e-209;  
 Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Y 1 TGCCGGGCTGCGGTGAGCCCTGATCCGCTCCGGGCTCCAGTGCCTGCTGCTGCTGA 60  
 b 1 TGCCGGGCTGCGGTGAGCCCTGATCCGCTCCGGGCTCCAGTGCCTGCTGCTGCTGA 60  
 Y 61 CGCTGCTCATCGCCGGCTCTACGGTGTGAGGCGCAGGATGCCCTGAAGTGCCTGGCG 120  
 b 61 CGCTGCTCATCGCCGGCTCTACGGTGTGAGGCGCAGGATGCCCTGAAGTGCCTGGCG 120  
 Y 121 CCCCTGCTACTCCCTGACTGTGACCCCTTTTCCTCTACTCCCTCCCTCCCTCCCTAG 180  
 b 121 CCCCTGCTACTCCCTGACTGTGACCCCTTTTCCTCTACTCCCTCCCTCCCTCCCTAG 180  
 Y 181 GATCCCTGAGCTTCAGATCTGGATGGCAGGATGGCTTCAGATGGCTGAAC 240  
 b 181 GATCCCTGAGCTTCAGATCTGGATGGCAGGATGGCTTCAGATGGCTGAAC 240  
 Y 241 CCTCCCGTATTTTAAACTGGTTCCTCGTGAGAGCTGTGAATCGGGCTCTGTATGC 300  
 b 241 CCTCCCGTATTTTAAACTGGTTCCTCGTGAGAGCTGTGAATCGGGCTCTGTATGC 300  
 Y 301 GCTTGAGAAAGCCCATTCATGAGGCGAAGCCAGTGGTCCCTCCCACTCCCGGACC 360  
 b 301 GCTTGAGAAAGCCCATTCATGAGGCGAAGCCAGTGGTCCCTCCCACTCCCGGACC 360  
 Y 361 CCCCTCTCCCAATGACAGCTCTCCCGGCTCATCCCCCCCCCCCCCGGTCGCCGCG 420  
 b 361 CCCCTCTCCCAATGACAGCTCTCCCGGCTCATCCCCCCCCCCCCCGGTCGCCGCG 420  
 Y 421 CTGCGGCACCTTCAGATCGATCTGGATGGCAGGATGGCTTCAGATGGCTGAAC 480  
 b 421 CTGCGGCACCTTCAGATCGATCTGGATGGCAGGATGGCTTCAGATGGCTGAAC 480  
 Y 481 CTGCGGCCTATTTTAAACTGGTTCCTCGTGAGAGCTGTGAATCGGGCTCTGTATGC 540  
 b 481 CTGCGGCCTATTTTAAACTGGTTCCTCGTGAGAGCTGTGAATCGGGCTCTGTATGC 540  
 Y 541 GCTTGAGAAAGCCCATTCATGAGGCGAAGCCAGTGGTTCCTCCCACTCCCGGACC 600  
 b 541 GCTTGAGAAAGCCCATTCATGAGGCGAAGCCAGTGGTTCCTCCCACTCCCGGACC 600  
 Y 601 CCCCTCTCCCAATGACAGCTCTCCCGGCTCATCCCCCCCCCCCCCGGTCGCCGCG 660  
 b 601 CCCCTCTCCCAATGACAGCTCTCCCGGCTCATCCCCCCCCCCCCCGGTCGCCGCG 660  
 Y 661 CTGCGGCACCTTCAGATCGATCTGGATGGCAGGATGGCTTCAGATGGCTGAAC 720  
 b 661 CTGCGGCACCTTCAGATCGATCTGGATGGCAGGATGGCTTCAGATGGCTGAAC 720  
 Y 721 AAGCCAGCTCTAGTGGATCCCGGGCTCAGATCTGTAGGCGCAGTAGTCCAGGGT 780  
 b 721 AAGCCAGCTCTAGTGGATCCCGGGCTCAGATCTGTAGGCGCAGTAGTCCAGGGT 780  
 Y 781 TCCTTGATGATGATCATATCTATCTGTCCCTTTTTCACAGCTCGGGGTGAGGACA 840  
 b 781 TCCTTGATGATGATCATATCTATCTGTCCCTTTTTCACAGCTCGGGGTGAGGACA 840  
 Y 841 AACTCTTCGGGCTCTTCCAGTGGGATTCAGCGGTATCGATTAAGCTTGATCTGTGAC 900  
 b 841 AACTCTTCGGGCTCTTCCAGTGGGATTCAGCGGTATCGATTAAGCTTGATCTGTGAC 900

QY 901 ATGCGGATCCCGTCTGTTTACACGTCGTGACTGGGAAACCC 944  
 Db 901 ATGCGGATCCCGTCTGTTTACACGTCGTGACTGGGAAACCC 944  
 RESULT 2  
 MUSPACOLL 32492 bp DNA linear ROD 09-JUN-1995  
 LOCUS Mouse pro-alpha (II) collagen chain gene, complete cds.  
 DEFINITION M65161  
 ACCESSION M65161.1 GI:854650  
 VERSION alpha-1 type II collagen; cartilage-specific collagen;  
 KEYWORDS fibril-forming collagen.  
 SOURCE Mouse (strain C57/Black) DNA.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 32492)  
 AUTHORS Metzaranta, M., Toman, D., de Crombrughe, B. and Vuorio, E.  
 TITLE Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and alternative splicing  
 J. Biol. Chem. 266 (25), 16862-16869 (1991)  
 MEDLINE 91358489  
 PUBMED 1885613  
 REFERENCE 2 (bases 1792 to 32492)  
 AUTHORS Vuorio, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-OCT-1991) E. Vuorio, Department of Medical  
 Biochemistry, University of Turku, Finland  
 REFERENCE 3 (bases 1 to 32492)  
 AUTHORS Vuorio, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-JAN-1995) E. Vuorio, Department of Medical  
 Biochemistry, University of Turku, Finland  
 COMMENT On Jun 9, 1995 this sequence version replaced gi:200213.  
 FEATURES  
 Location/Qualifiers  
 1..32492  
 /organism="Mus musculus"  
 /strain="C57/Black"  
 /db\_xref="taxon:10090"  
 1792..2893  
 /gene="pro-alpha (II) collagen"  
 /evidence=experimental  
 2866..2870  
 /gene="pro-alpha (II) collagen"  
 /evidence=experimental  
 TATA\_signal  
 mRNA  
 join(2894..3130,6930..7136,8245..8261,8649..8681,  
 8935..8988,10144..10221,10841..10885,10991..11044,  
 11375..11428,12118..12171,12462..12515,12610..12663,  
 12943..12987,13461..13514,16128..16172,16551..16604,  
 18100..18198,18490..18534,18622..18720,18866..18919,  
 19291..19398,19759..19812,19897..19995,20134..20187,  
 20653..20751,21131..21184,21575..21628,21875..21928,  
 22157..22210,22444..22488,22675..22773,23025..23132,  
 23471..23524,23780..23833,24195..24248,24530..24583,  
 24867..24974,25446..25499,25935..25988,26616..26777,  
 26947..27054,27249..27356,27574..27627,27763..27870,  
 28061..28114,28227..28334,28569..28622,29055..29162,  
 29448..29736,30089..30276,30574..30816,31273..32492)  
 /partial  
 /gene="pro-alpha (II) collagen"  
 mRNA  
 join(2894..3130,6930..7136,8245..8261,8649..8681,  
 9130..9234,10144..10221,10841..10885,10991..11044,  
 11375..11428,12118..12171,12462..12515,12610..12663,  
 12943..12987,13461..13514,16128..16172,16551..16604,  
 18100..18198,18490..18534,18622..18720,18866..18919,  
 19291..19398,19759..19812,19897..19995,20134..20187,  
 20653..20751,21131..21184,21575..21628,21875..21928,  
 22157..22210,22444..22488,22675..22773,23025..23132,  
 23471..23524,23780..23833,24195..24248,24530..24583,  
 24867..24974,25446..25499,25935..25988,26616..26777,  
 26947..27054,27249..27356,27574..27627,27763..27870,  
 28061..28114,28227..28334,28569..28622,29055..29162,  
 29448..29736,30089..30276,30574..30816,31273..32492)



530 GCTCTGATGCGTTGAGAAAGCCCATTCATGAGAGGCAAGGCCAGTGGTCCCCCA 589  
 5092 GCTCTGATGCGTTGAGAAAGCCCATTCATGAGAGGCAAGGCCAGTGGTCCCCCA 5151  
 590 ACTCCGAGCCCTCTCCACATGACAGCTCCCGCCTCATCCCTCCCTCCCTCC 649  
 5152 ACTCCGAGCCCTCTCTTACATGACAGCTCCCGCCTCATCCCTCCCTCCCTCC 5211  
 650 CCGGTGCGCGCTCCCGCACCTCCAGATCTCCAGTAGAGGATCTGGAC 701  
 5212 CCGGTGCGCGCTCCCGCACCTCCCGGCTCCAGCCCGCGCAGAGCGG 5263

RESULT 3  
 AC113444 183204 bp DNA linear HTG 02-JUN-2002  
 Mus musculus clone RP23-268C9, WORKING DRAFT SEQUENCE, 21 ordered  
 pieces.  
 AC113444  
 AC113444.2 GI-21313781  
 HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 183204)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP23-268C9  
 2 (bases 1 to 183204)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhvalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choepel, V., Colangelo, M., Collins, S., Collumore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kanat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,  
 Landers, T., Lehoczyk, J., Levine, R., Liu, G., Maclean, C.,  
 Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,  
 McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,  
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,  
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
 JOURNAL  
 Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 183204)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,  
 Boukhvalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
 Chazaro, B., Choepel, V., Colangelo, M., Collins, S., Collumore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N., Hagos, B.,  
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A.,  
 Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T.,  
 Lehoczyk, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J.,  
 Marquis, N., Matthews, C., McCarthy, M., McKernan, K., Meldrim, J.,  
 Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,

Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S.,  
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
 Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-JUN-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 2, 2002 this sequence version replaced gi:19033610.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIER

Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L20375

Center clone name: 268 C 9

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 176089 bases at least Q40

Consensus quality: 178979 bases at least Q30

Consensus quality: 180313 bases at least Q20

Insert size: 183000; agarose-fp

Insert size: 18204; sum-of-contigs

Quality coverage: 6.2 in Q20 bases; agarose-fp

Quality coverage: 6.2 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 21 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 751: contig of 751 bp in length

\* 752 851: gap of 100 bp

\* 852 2018: contig of 1167 bp in length

\* 2019 2118: gap of 100 bp

\* 2119 3480: contig of 1362 bp in length

\* 3481 3580: gap of 100 bp

\* 3581 5074: contig of 1494 bp in length

\* 5075 5174: gap of 100 bp

\* 5175 6381: contig of 1207 bp in length

\* 6382 6481: gap of 100 bp

\* 6482 7372: contig of 891 bp in length

\* 7373 7472: gap of 100 bp

\* 7473 9254: contig of 1782 bp in length

\* 9255 9354: gap of 100 bp

\* 9355 11356: contig of 2002 bp in length

\* 11357 11456: gap of 100 bp

\* 11457 13778: contig of 2322 bp in length

\* 13779 13878: gap of 100 bp

\* 13879 15543: contig of 1655 bp in length

\* 15544 15643: gap of 100 bp

\* 15644 18289: contig of 2646 bp in length

\* 18290 18389: gap of 100 bp

\* 18390 22304: contig of 3915 bp in length

\* 22305 22404: gap of 100 bp

\* 22405 27545: contig of 5141 bp in length

\* 27546 27645: gap of 100 bp

\* 27646 34213: contig of 6568 bp in length

\* 34214 34313: gap of 100 bp

\* 34314 42078: contig of 7765 bp in length

\* 42079 42178: gap of 100 bp

\* 42179 58590: contig of 16412 bp in length



\* 58591 58690: gap of 100 bp  
\* 58691 74191: contig of 15501 bp in length  
\* 74192 74291: gap of 100 bp  
\* 74292 95692: contig of 21401 bp in length  
\* 95693 95792: gap of 100 bp  
\* 95793 113790: contig of 17998 bp in length  
\* 113791 113890: gap of 100 bp  
\* 113891 137701: contig of 23811 bp in length  
\* 137702 137801: gap of 100 bp  
\* 137802 183204: contig of 45403 bp in length.

## FEATURES

Location/Qualifiers  
1..183204  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-268C9"  
/clone\_lib="RPCI-23 Female Mouse BAC"

## misc\_feature

1..751  
/note="assembly\_fragment"  
852..2018  
/note="assembly\_fragment"  
2119..3480  
/note="assembly\_fragment"  
3581..5074  
/note="assembly\_fragment"  
5175..6381  
/note="assembly\_fragment"  
6482..7372  
/note="assembly\_fragment"  
7473..9254  
/note="assembly\_fragment"  
9355..11356  
/note="assembly\_fragment"  
11457..13778  
/note="assembly\_fragment"  
13879..15543  
/note="assembly\_fragment"  
15644..18289  
/note="assembly\_fragment"  
18390..22304  
/note="assembly\_fragment"  
22405..27545  
/note="assembly\_fragment"  
27646..34213  
/note="assembly\_fragment"  
34314..42078  
/note="assembly\_fragment"  
42179..58590  
/note="assembly\_fragment"  
58691..74191  
/note="assembly\_fragment"  
74292..95692  
/note="assembly\_fragment"  
95793..113790  
/note="assembly\_fragment"  
113891..137701  
/note="assembly\_fragment"  
137802..183204  
/note="assembly\_fragment"  
clone end:77  
vector side:right"

SE COUNT 48115 a 43085 c 42967 g 47055 t 2002 others  
IGIN

Query Match 24.3%; Score 229.2; DB 2; Length 183204;  
Best Local Similarity 96.7%; Pred. No. 2.6e-43;  
Matches 234; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

195 TTGCAGATCTGGGATTCGACGATGCTCCAGATGGGCTGAACCTCCCGATTATTA 254

151633 TTTCGGTTGGGGGATTCGACGATGCTCCAGATGGGCTGAACCTCCCGATTATTA 151692

255 TTAAACTGGTTCCTCGTGGAGAGCTGTGAATCGGCTCTGTATGGCTTGAGAAAGCC 314

|||||

Db 151693 TTAAACTGGTTCCTCGTGGAGAGCTGTGAATCGGCTCTGTATGGCTTGAGAAAGCC 151752  
QY 315 CCATTATGAGAGGCAAGGCCAGTGGGTCCCACTCCCGACCCCTCTCCACAA 374  
Db 151753 CCATTATGAGAGGCAAGGCCAGTGGGTCCCACTCCCGACCCCTCTCCACAA 151812  
QY 375 TGCACAGCTCCCGCCCTCATCCCCCCCCACCCCGTCCGCGCTGCGCCACCTTC 434  
Db 151813 TGCACAGCTCCCGCCCTCATCCCCCCCCACCCCGTCCGCGCTGCGCCACCTTC 151872  
QY 435 AG 436  
Db 151873 CG 151874

## RESULT 4

AX114853 AX114853 7175 bp DNA linear PAT 11-MAY-2001  
LOCUS Sequence 13 from Patent WO0129208.  
DEFINITION  
ACCESSION AX114853  
VERSION AX114853.1 GI:14031795

## KEYWORDS

synthetic construct.

## SOURCE

synthetic construct  
artificial sequences

## ORGANISM

1 (bases 1 to 7175)

## REFERENCE

Kuehn, R., von Melchner, H. and Altschmied, J.

## AUTHORS

Conditional gene trapping construct for the disruption of genes

## TITLE

Patent: WO 0129208-A 13 26-APR-2001.

## JOURNAL

ARTEMIS Pharmaceuticals GmbH (DE); Frankgen Biotechnologie AG (DE)

## FEATURES

Location/Qualifiers  
1..7175  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="vector PRK57SA-beta"

## BASE COUNT

1614 a 1892 c 2002 g 1667 t

## ORIGIN

Query Match 22.9%; Score 216.6; DB 6; Length 7175;  
Best Local Similarity 98.2%; Pred. No. 3.2e-40;  
Matches 219; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 722 AGCCAAGCTCTAGTGGATCCCGGGCTGCAGATCTGTAGGGCGCAGTAGTCCAGGGTTT 781

Db 488 AGGCTAGAACTAGTGGATCCCGGGCTGCAGATCTGTAGGGCGCAGTAGTCCAGGGTTT 547

QY 782 CCTTGATGATGTCATATTATCTGTCCCTTTTTTTTCCACAGCTCGCGGTGAGGACAA 841

Db 548 CCTTGATGATGTCATATTATCTGTCCCTTTTTTTTCCACAGCTCGCGGTGAGGACAA 607

QY 842 ACTCTTCGGGCTCTTCCAGTGGGATCGACGATCGATTAAGCTTGATGATCTGTGACA 901

Db 608 ACTCTTCGGGCTCTTCCAGTGGGATCGACGATCGATTAAGCTTGATGATCTGTGACA 667

QY 902 TGGCGGATCCCGTGGTTTTTACAAAGTCTGTGACTGGGAAAACCC 944

Db 668 TGGCGGATCCCGTGGTTTTTACAAAGTCTGTGACTGGGAAAACCC 710

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Rhodes, C.S. and Yamada, Y.  
Identification of a minimum enhancer sequence for the type II  
collagen gene reveals several core sequence motifs in common with  
the link protein gene  
J. Biol. Chem. 271 (8), 4298-4303 (1996)  
96224009  
8626777

Location/Qualifiers  
1. .823  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/name="Col2A1"  
/notes="collagen type II"  
/name="Col2A1"  
/number=1  
165 a 258 c 211 g 189 t

Query Match 18.3%; Score 172.8; DB 10; Length 823;  
Best Local Similarity 81.5%; Pred. No. 6.5e-30;  
Matches 212; Conservative 0; Mismatches 47; Indels 1; Gaps 1;  
Y 410 CCCGTGCGCCGCTGCGCCACCTTCAGATCGATCTGGGATTGCGAGCGATGCTCCAGA 469  
b 243 CTTGAGTCTCTTTGAGGCTGTTGGTGTAGGATTGCGAGCGATGCTCCAGA 302  
Y 470 TGGGCTGAACCCCTGCGCGCTATTTATTTAACTGGTTCCTCGTGGAGAGCTGTGAATCGG 529  
b 303 TGGGCTGAACCCCTGCGCGCTATTTATTTAACTGGTTCCTCGTGGAGAGCTGTGAATCGG 362  
Y 530 GCTCTGTATGGCTTGAGAAAGCCCATTCATGAGAGCGAGCCGCTGGT-CCCCC 588  
b 363 GCTCTGTATGGCTTGAGAAAGCCCATTCATGAGAGCGAGCCGCTGGT-CCCCC 422  
Y 589 AACTCCCGGACCCCTCTCCCAATGACACGCTCCCGCCCTCATCCCCCCCCCCAC 648  
b 423 GACTCCCGGACCCCTCTCCCAATGACATCCCTCCCTGTGCGCGCTGCGCCAC 482  
Y 649 CCCCCTGCGCGCTGCGCGC 668  
b 483 CTCCCGGCTGCGCGCCCGC 502

RESULT 6  
AC098511  
AC098511  
AC098511.3 GI:21729588  
HTG; HTGS PHASE1.  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 197631)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C.,  
Albrooke, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T.,  
Barbieri, J., Benton, J., Simge, K., Blankenburg, K., Bonnin, D.,  
Buck, J., Bowie, S., Brieva, N., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burrell, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flaggs, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseg, H.,  
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwundu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shoaib, N., Sisson, I.,  
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 197631)  
Worley, K.C.  
Direct Submission  
Submitted (24-OCT-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 197631)  
Worley, K.C.  
Direct Submission  
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 11, 2002 this sequence version replaced gi:17973150.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GHFR  
Center clone name: CH230-66N4  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 129644 bases at least Q40  
Consensus quality: 135710 bases at least Q30  
Consensus quality: 140703 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 73 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1254: contig of 1254 bp in length  
\* 1255 1354: gap of unknown length  
\* 1355 2715: contig of 1361 bp in length  
\* 2716 2815: gap of unknown length  
\* 2816 4418: contig of 1603 bp in length  
\* 4419 4518: gap of unknown length

\* 4519 5912: contig of 1394 bp in length  
\* 5913 6012: gap of unknown length  
\* 6013 7052: contig of 1040 bp in length  
\* 7053 7152: gap of unknown length  
\* 7153 8253: contig of 1101 bp in length  
\* 8254 8353: gap of unknown length  
\* 8354 9424: contig of 1071 bp in length  
\* 9425 9524: gap of unknown length  
\* 9525 10883: contig of 1359 bp in length  
\* 10884 12607: gap of unknown length  
\* 12608 12607: contig of 1624 bp in length  
\* 12609 14171: gap of unknown length  
\* 12708 14171: contig of 1464 bp in length  
\* 14172 14271: gap of unknown length  
\* 14272 15778: contig of 1507 bp in length  
\* 15779 15878: gap of unknown length  
\* 15879 17195: contig of 1317 bp in length  
\* 17196 17295: gap of unknown length  
\* 17296 18822: contig of 1527 bp in length  
\* 18823 18923: gap of unknown length  
\* 18924 19943: contig of 1027 bp in length  
\* 19944 20043: gap of unknown length  
\* 20044 21278: contig of 1229 bp in length  
\* 21279 21378: gap of unknown length  
\* 21379 22428: contig of 1050 bp in length  
\* 22429 23793: gap of unknown length  
\* 23794 23893: gap of unknown length  
\* 23894 25052: contig of 1159 bp in length  
\* 25053 25152: gap of unknown length  
\* 25153 26328: contig of 1176 bp in length  
\* 26329 26428: gap of unknown length  
\* 26429 27957: contig of 1529 bp in length  
\* 27958 28057: gap of unknown length  
\* 28059 29478: contig of 1422 bp in length  
\* 29479 31146: contig of 1567 bp in length  
\* 31147 31247: gap of unknown length  
\* 31248 32886: gap of unknown length  
\* 32887 34341: contig of 1455 bp in length  
\* 34342 34441: gap of unknown length  
\* 34442 35975: contig of 1534 bp in length  
\* 35976 36075: gap of unknown length  
\* 36076 37371: gap of unknown length  
\* 37372 39561: contig of 2091 bp in length  
\* 39562 39661: gap of unknown length  
\* 39662 42023: contig of 2368 bp in length  
\* 42024 42123: gap of unknown length  
\* 42124 43520: contig of 1391 bp in length  
\* 43521 44754: contig of 1134 bp in length  
\* 44755 44854: gap of unknown length  
\* 44855 46992: contig of 2138 bp in length  
\* 46993 47092: gap of unknown length  
\* 47093 49984: contig of 1892 bp in length  
\* 49985 50415: gap of unknown length  
\* 50416 50515: gap of unknown length  
\* 50516 52767: contig of 2252 bp in length  
\* 52768 52867: gap of unknown length  
\* 52869 55049: contig of 3082 bp in length  
\* 55050 58189: contig of 2140 bp in length  
\* 58190 58289: gap of unknown length  
\* 58290 60246: contig of 1957 bp in length  
\* 60247 60346: gap of unknown length  
\* 60347 62936: contig of 2589 bp in length  
\* 62937 65056: contig of 3021 bp in length  
\* 65057 66156: gap of unknown length  
\* 66157 69116: contig of 2960 bp in length

\* 59117 59216: gap of unknown length  
\* 59217 70782: contig of 1566 bp in length  
\* 70783 70882: gap of unknown length  
\* 70883 74187: contig of 3305 bp in length  
\* 74188 74287: gap of unknown length  
\* 74289 77957: contig of 3670 bp in length  
\* 77958 78057: gap of unknown length  
\* 78059 81067: contig of 3010 bp in length  
\* 81068 81167: gap of unknown length  
\* 81169 83664: contig of 2497 bp in length  
\* 83665 83764: gap of unknown length  
\* 83765 86822: contig of 3058 bp in length  
\* 86823 86922: gap of unknown length  
\* 86923 89801: contig of 2879 bp in length  
\* 89802 89901: gap of unknown length  
\* 89902 94055: contig of 4154 bp in length  
\* 94056 94156: gap of unknown length  
\* 94157 97058: gap of unknown length  
\* 97059 100680: contig of 3523 bp in length  
\* 100681 100780: gap of unknown length  
\* 100781 104720: contig of 3940 bp in length  
\* 104721 104820: gap of unknown length  
\* 104821 107723: contig of 2903 bp in length  
\* 107724 107823: gap of unknown length  
\* 107824 110257: contig of 2434 bp in length  
\* 110258 110357: gap of unknown length  
\* 110359 115026: contig of 4669 bp in length

Query Match 18.1%; Score 171; DB 2; Length 197631;  
Best Local Similarity 86.6%; Pred. No. 9.7e-30;  
Matches 200; Conservative 0; Mismatches 30; Indels 1; Gaps 1;

QY 439 CGATCTGGATTGCGACGATGCTCCAGATGGCTGAACCTCCCGTATTATTATTA 498  
Db 97188 GTTGAGGATTGCGACGATGCTCCAGATGGCTGAACCTCCCGTATTATTATTA 97247

QY 499 AACTGGTTCCTCTGGAGAGCTGTGAATCGGGCTCTGTATGGCTTGAGAAAAGCCCAT 558  
Db 97248 AACTGGTTCCTCTGGAGAGCTGTGAATCGGGCTCTGTATGCACTCGAGAAAAGCCCAT 97307

QY 559 TCATGAGAGGAGGAGGAGGAGTGGT-CCCCCACTCCCGAGCCCGCTCCCAATGC 617  
Db 97308 TCATGAGAGGAGGAGGAGTGGTCCCGGAGCTCCCGAGCCCGCTCCCAATGC 97367

QY 618 ACAGCTCTCCCGCCTCATCCCCCCCCCACCACCCCGCTCCCGCTCCCGC 668  
Db 97368 ATCCCCCTCCCTGTGCGCGCTGCGGACCTCCCGGCTCCGGCCCCGC 97418

RESULT 7  
MUSCOL2A1A MUSCOL2A1A 861 bp DNA linear ROD 10-AUG-2001  
LOCUS Mus musculus alpha-1 type II collagen (Col2a-1) gene, exon 1 and  
DEFINITION partial cds.  
ACCESSION M63710  
VERSION M63710.1 GI:192661  
KEYWORDS Mus musculus.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 861)  
AUTHORS Cheah, K.S., Au, P.K., Lau, E.T., Little, P.F. and Stubbs, L.  
TITLE The mouse Col2a-1 gene is highly conserved and is linked to Int-1  
on chromosome 15  
JOURNAL Mamm. Genome 1 (3), 171-183 (1991)  
MEDLINE 92182509  
PUBMED 1797232  
FEATURES Location/Qualifiers  
source 1. .861  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

```

gene
CAAT_signal
GC_signal
repeat_region
repeat_region
TATA_signal
repeat_region
mRNA
exon
5'UTR
CDS
sig_peptide
mat_peptide
repeat_region
ASE COUNT 144 a 268 c 280 g 169 t
RIGIN

Query Match 18.1%; Score 170.6; DB 10; Length 861;
Best Local Similarity 97.7%; Pred. No. 2.1e-29;
Matches 173; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

y 4 CGGCTCGGGTGAGCCCTGATCGGCTCGGGCTCCCGAGTCGCTGGTGTGCTGAGCG 63
b 389 CAGGCTCGGGTGAGCCATGATCGGCTCGGGCTCCCGAGTCGCTGGTGTGCTGAGCG 448
y 64 TGCTCATCGCGGCTCCTACGGTGTGAGGCGCAGGATCGCGTAACTGCGCGCGGCC 123
b 449 TGCTCATCGCGGCTCCTACGGTGTGAGGCGCAGGATCGCGTAACTGCGCGCGGCC 508
y 124 CTGCTACTTCCCTGACTGTGACCCCTTTCTCCCTACTCCCTCCCGCAAGTACTAG 180
b 509 CTGCTACTTCCCTGACTGTGACCCCTTTCTCCCTACTCCCTCCCGCAAGTACTAG 565

RESULT 8
LOCUS RATCG2A1A 1404 bp DNA linear ROD 27-APR-1993
DEFINITION Rat alpha-1 procollagen(II) gene, promoter region and exon 1.
ACCESSION M10613
VERSION M10613.1 GI:203426
KEYWORDS alpha-1 type II collagen.
SOURCE Rat DNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

55..>861
/gene="Col2a-1"
55..58
/gene="Col2a-1"
/evidence=experimental
59..137
/gene="Col2a-1"
71..107
/rpt_type=direct
186..196
/evidence=experimental
227..232
/rpt_type=direct
242..273
/evidence=experimental
255..>490
/rpt_type=direct
255..>490
/product="alpha-1 type II collagen"
255..490
/gene="Col2a-1"
255..405
/gene="Col2a-1"
406..>490
/gene="Col2a-1"
/codon_start=1
/product="alpha-1 type II collagen"
/protein_id="AAA37435.1"
/db_xref="GI:553895"
/translation="MIRLGAPQSLVLLTLLIAAIVLRCQGQDA"
406..468
/gene="Col2a-1"
469..>490
/gene="Col2a-1"
/product="alpha-1 type II collagen"
812..851
/notes="sp1 binding site"
/evidence=experimental
/rpt_type=direct
ASE COUNT 144 a 268 c 280 g 169 t
RIGIN

55..>861
/gene="Col2a-1"
55..58
/gene="Col2a-1"
/evidence=experimental
59..137
/gene="Col2a-1"
71..107
/rpt_type=direct
186..196
/evidence=experimental
227..232
/rpt_type=direct
242..273
/evidence=experimental
255..>490
/rpt_type=direct
255..>490
/product="alpha-1 type II collagen"
255..490
/gene="Col2a-1"
255..405
/gene="Col2a-1"
406..>490
/gene="Col2a-1"
/codon_start=1
/product="alpha-1 type II collagen"
/protein_id="AAA40920.1"
/db_xref="GI:554430"
/translation="MIRLGAPQSLVLLTLLIAIVLRCQGQDA"
1237..>1404
/gene="Col2a1"
/number=1
BASE COUNT 228 a 469 c 407 g 300 t
ORIGIN

Query Match 14.7%; Score 138.6; DB 10; Length 1404;
Best Local Similarity 88.8%; Pred. No. 5.8e-22;
Matches 150; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 TGCGCGCCCTCGCGGTGAGCCCTGATCCGCTCGGGGCTCCCGAGTCGCTGGTGTGCTGA 60
DB 1132 TGCGCGCCCTCGCGGTGAGCCATGATCCGCTCGGGGCTCCCGAGTCGCTGGTGTGCTGA 1191
QY 61 CGTGCTCATCGCGCGGCTCTACGGTGTGAGGCGCAGAGTGCCTGTAAGTCGCCCGCGCG 120
DB 1192 CGTGCTCATCGCGCGGCTCTACAAATGTGAGGCGCAGAGTGCCTGTAAGTCGCCCGCGCG 1251
QY 121 CCCTCGCTACTTCCCTGACTGTGACCCCTTTCTCCCTACTTCCCTCCCTCC 169
DB 1252 CCCTCGCTACTTCCCTGACTGTGACCCCTTTCTCCCTACTTCCCTCCCTCC 1300

RESULT 9
HUMCOLII
LOCUS Human alpha-1 collagen type II gene, exons 1, 2 and 3.
DEFINITION
ACCESSION M60299
VERSION M60299.1 GI:180883
KEYWORDS alpha-1 type II collagen.
SOURCE Human DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 6723)
The human type II procollagen gene: identification of an additional
protein-coding domain and location of potential regulatory
sequences in the promoter and first intron
Genomics 8 (1), 41-48 (1990)
91184811
MEDLINE
PUBMED 2081599
REFERENCE 2 (bases 1 to 6723)
AUTHORS Adams,M.E., Huang,D.Q., Yao,L.Y. and Sandell,L.J.
TITLE Extraction and isolation of mRNA from adult articular cartilage
JOURNAL Anal. Biochem. 202 (1), 89-95 (1992)
MEDLINE 92321501
PUBMED 1621990
FEATURES
Location/Qualifiers
1..6723
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/map="12q13"
/tissue_lib="of T.manlatis"
667. .6723
/genes="COL2A1"
667. .672
/genes="COL2A1"
/notes="G00-119-063"
join(852. .936,5010. .5216,6707. .6723)
/partial
/genes="COL2A1"
/codon_start=1
/product="alpha-1 type II collagen"
/protein_id="AAA73873.1"
/db_xref="GI:180884"
/db_xref="GBB:G00-119-063"
/translation="MIRLGAQPSLVLLTLVAUVRCQQDVQAGSCVQDQGRYNDK
DVKKPEPCRICVDGTGVLCDIICEDVKDCLSPFEGECPCPTDLATASQPGP
K"
937. .5009
/genes="COL2A1"
/notes="G00-119-063"
/number=1
5010. .5216
/genes="COL2A1"
/notes="G00-119-063"
/number=2
5217. .6706
/genes="COL2A1"
/notes="G00-119-063"
/number=2
SE COUNT 1357 a 1865 c 1820 g 1681 t
IGIN
Query Match 13.2%; Score 124.6; DB 9; Length 6723;
Best Local Similarity 78.4%; Pred. No. 9e-19; Indels 2; Gaps 2;
Matches 174; Conservative 0; Mismatches 46;
445 GGGATTGGCAGCGATGCTCCAGAT-GGGCTGAAACCTCGCCGCTATTATTAACTG 503
2966 GAGATTGGCAGCGATGCTCCAGATGGGCTGAAAGCTGCCGCTATTATTAACTG 3025
504 GTTCTCTGTGAGAGCTGTGATCGGCTCTGTATGGCTTGAGAAAGCCCATCATG 563
3026 GTTCTCTGAGAGACCTGTGAATCGGCTCTGTGTGGCTCGAGAAAGCCCATCATG 3085
564 AGAGCAAGGCCAGTGGG-TCCCCAACTCCCGACCCCTCTCCCAATGACAGC 622
3086 AGACAGGCTGCAAGTGGTCTCTGCTACTCCAGACCCCTCTCCACATGCCCCC 3145
623 CTCCTCCGCTCATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 664
3146 TGTGCCAGCGCCGCGCTCTCTCGCTCCAGCCCTGGCCAG 3187

```

```

SULT 10
CUS AC004801 193561 bp DNA linear PRI 02-FEB-1999
FINITION Homo sapiens 12q13.1 PAC RPC11-228P16 (Roswell Park Cancer
Institute Human PAC Library) complete sequence.
CESSION AC004801
RSION AC004801.1 GI:4204244
YWORDS HTG.
URCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193561)
Muzny, D., Arenson, A. D., Bouck, J., Bunac, C., Chen, Z., Ding, Y.,
Dugan, S., Durbin, J., Forcum, J., Garcia, C., Gorrell, J. H.,
Gorrell, L. L., Hernandez, J., Issar, A., Jackson, L., Kneitz, S.,
Kondejewski, N., Lau, S., Leal, B., Lee, E., Lichtarge, O., Liu, W.,
Logan, O., Lu, J., Marondel, I., Martinez, C., Merscher, S., Miller, A.,
Montgomery, K., Oswal, G., Pampell, L. R., Parish, B. J., Perez, L.,

```

Rashid, N. D., Rives, C., Scherer, S. E., Shen, H., Shim, C., Simon, M.,  
Vo, Q., Williamson, A., Worley, K. C., Xiang, A. M., Yang, R., Yu, W.,  
Zhou, X., Kucherlapati, R., Nelson, D. and Gibbs, R. A.

Unpublished  
Direct Submission

2 (bases 1 to 193561)  
Worley, K. C.

Direct Submission  
Submitted (06-JUN-1998) Molecular and Human Genetics, Baylor

College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 193561)  
Worley, K. C.

Direct Submission  
Submitted (30-JAN-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 193561)  
Worley, K. C.

Direct Submission  
Submitted (02-FEB-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jan 30, 1999 this sequence version replaced gi:3763910.  
INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email  
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.

FEATURES	Location/Qualifiers
source	1. .193561 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="12q13.1" /clone="RPC11-228P16" complement(123. .408) /rpt_family="AluSg" complement(805. .895) /rpt_family="MIR" complement(join(1336. .1461,3799. .4239,4353. .4461, 4906. .4950,5702. .5867,6125. .6314,7789. .7866,8746. .8819, 16323. .16415,19776. .19839)) /gene="Phosphofructokinase U24183" 1837. .2124 /rpt_family="AluSx" 2129. .2430 /rpt_family="AluSg"
repeat_region	
repeat_region	
gene	
repeat_region	
repeat_region	

```
repeat_region 2432..2995
/rpt_family="L1MD1"
repeat_region 2996..3293
/rpt_family="AluY"
repeat_region 3358..3533
/rpt_family="L1MD1"
STS 5796..6186
/standard_name="MS9723"
repeat_region complement(9157..9270)
/rpt_family="MIR"
repeat_region 9822..9954
/rpt_family="FLAM_C"
complement(10598..10586)
repeat_region /rpt_family="L2"
10779..10908
/rpt_family="FLAM_A"
repeat_region complement(11422..11512)
/rpt_family="L2"
repeat_region complement(11865..12121)
/rpt_family="L2"
13424..13538
/rpt_family="FLAM_A"
repeat_region complement(13542..13842)
/rpt_family="AluSp"
repeat_region complement(15183..15241)
/rpt_family="MIR"
repeat_region complement(18098..18312)
/rpt_family="AluX"
repeat_region complement(19370..18662)
/rpt_family="MER58B"
18679..18824
/rpt_family="L2"
repeat_region complement(20222..20319)
/rpt_family="L2"
20746..20904
/rpt_family="L1MC2"
repeat_region 20905..21430
/rpt_family="MER4B"
repeat_region 21504..22403
/rpt_family="L1MC2"
repeat_region 22404..22705
/rpt_family="AluJb"
repeat_region 22711..22883
/rpt_family="L1M4"
repeat_region 22945..23061
/rpt_family="(GGA)n"
repeat_region 23071..23380
/rpt_family="L1M4"
repeat_region complement(23384..23505)
/rpt_family="MER4D"
repeat_region 23506..23927
/rpt_family="LTR7"
repeat_region 23957..24086
/rpt_family="HERVH"
repeat_region 24147..24891
/rpt_family="HERVH"
repeat_region 25227..26737
/rpt_family="HERVH"
repeat_region 26724..27571
/rpt_family="HERVH"
repeat_region 27558..27838
/rpt_family="HERVH"
repeat_region 28743..29303
/rpt_family="L2"
repeat_region 29790..29878
/rpt_family="(GGA)n"
repeat_region 30042..30166
/rpt_family="L2"
repeat_region complement(32453..32481)
/rpt_family="(CAAA)n"
repeat_region 34536..34586
/rpt_family="L2"
repeat_region complement(34879..34911)
```

```
repeat_region /rpt_family="(GA)n"
complement(34919..35215)
/rpt_family="AluSg"
35770..35874
/rpt_family="MIR"
repeat_region complement(36315..36377)
/rpt_family="AluY"
repeat_region complement(36694..37001)
/rpt_family="AluX"
repeat_region complement(37839..38140)
/rpt_family="AluX"
repeat_region complement(38240..38464)
/rpt_family="MIR"
repeat_region complement(38490..38795)
/rpt_family="AluX"
repeat_region complement(39321..39609)
/rpt_family="AluX"
repeat_region complement(39799..40104)
/rpt_family="AluJb"
repeat_region complement(40313..40482)
/rpt_family="AluJo"
repeat_region complement(40484..40781)
/rpt_family="AluSg"
repeat_region complement(40784..40924)
/rpt_family="AluJo"
41522..41619
/rpt_family="MERSA"
repeat_region complement(41716..41903)
/rpt_family="MER20"
repeat_region complement(42195..42320)
/rpt_family="AluJo"
42389..42683
/rpt_family="AluSg"
repeat_region complement(43589..43888)
/rpt_family="AluX"
repeat_region complement(44447..44746)
/rpt_family="AluX"
```

Query Match 13.1%; Score 124; DB 9; Length 193561;

Best Local Similarity 80.2%; Pred No. 8,8e-19;

Matches 170; Conservative 0; Mismatches 40; Indels 2; Gaps 2;

```
QY 445 GGGATTGGCAGCGATGGCTTCCAGAT-GGGCTGAACCCCTGCCCTATTTTAACTG 503
Db 136781 GAGATTGGCAGCGATGGCTTCCAGATGGGCTGAACCGTGGCTATTTTAACTG 136840

QY 504 GTTCTCTGGAGAGCTGTGAATCGGCTCTGTATCGCTTGAGAAAAGCCCATTCATG 563
Db 136841 GTTCTCTGGAGAGCTGTGAATCGGCTCTGTATCGCTTGAGAAAAGCCCATTCATG 136900

QY 564 AGAGGCAAGGCCAGTGGG-TCCGCCAACTCCCGACCCCTCTCCCAATGCACAGC 622
Db 136901 AGAGGCAAGTCCAGTGGGTCTCTCGTACTCCAGACCCCTCTCCCAATGCACCC 136960

QY 623 CTCCCGCCCTCATCCCGCCCGCCACCTCTCGCTCCAG 654
Db 136961 TGTCCCGCCCGCCCGCCACCTCTCGCTCCAG 136992
```

## RESULT 11

```
LOCUS HSPOCOB1 5887 bp DNA linear PRI 14-AUG-1995
DEFINITION Human type II procollagen gene exon 1.
ACCESSION X58709 S40537
VERSION X58709.1 Gi:35659
KEYWORDS extracellular matrix protein; procollagen gene type II.
SOURCE Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5887)
AUTHORS Vakkula,M.S.
TITLE Direct Submission
```

JOURNAL Submitted (03-APR-1991) M.S. Vikkula, National Public Health Inst,  
Lab of Mol Genetics, Memmerheimintie 166, 00300 Helsinki, FINLAND  
REFERENCE 2 (bases 1 to 5887)  
AUTHORS Vikkula,M., Metsaranta,M., Syvanen,A.C., Ala-Kokko,L., Vuorio,E.  
and Peltonen,L.  
TITLE Structural analysis of the regulatory elements of the type-II  
procollagen gene. Conservation of promoter and first intron  
sequences between human and mouse  
JOURNAL Biochem. J. 285 (Pt 1), 287-294 (1992)  
MEDLINE 92344585  
PUBMED 1637314  
FEATURES  
source Location/Qualifiers  
1..5887  
/organism="Homo sapiens"  
/isolate="OA-patient"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="2,15,17,20,22,48,88,94"  
1190..1196  
/note="CCCTCC-binding sequence"  
1238..1244  
/note="CCCTCC-binding sequence"  
complement(1243..1248)  
/bound\_moiety="Sp1"  
complement(1309..1314)  
/bound\_moiety="Sp1"  
1346..1351  
/bound\_moiety="Sp1"  
1425..1430  
/bound\_moiety="Sp1"  
1429..1434  
/bound\_moiety="Sp1"  
1463..1468  
/bound\_moiety="Sp1"  
1472..1479  
/note="viral core enhancer"  
1514..1519  
1698..1782  
/note="type II procollagen gene exon 1"  
2411..2416  
/note="LF-A1 motif"  
3587  
/note="polymorphism (G/A)"  
3616..3621  
/note="ATF motif"  
3734..3739  
/note="LF-A1 motif"  
3804..3809  
/note="b-interferon silencer B"  
4116..4121  
/note="b-interferon silencer B"  
4254..4259  
/note="b-interferon silencer B"  
4973  
/note="polymorphism (C/A)"  
SE COUNT 1139 a 1763 c 1600 g 1385 t  
IGIN  
Query Match 13.0%; Score 122.4; DB 9; Length 5887;  
Best Local Similarity 79.7%; Pred. No. 3e-18;  
Matches 169; Conservative 0; Mismatches 41; Indels 2; Gaps 2;  
445 GGGATTGACGATGGTTCCAGT-GGGCTGAACCTGCCGTTATTTTAAACTG 503  
3830 GAGATTGACGATGGTTCCAGTGGGCTGAACCTGCCGTTATTTTAAACTG 3889  
504 GTTCTCTGTCGAGAGCTGTAATCGGCTCTGTATGCTTGAGAAAGCCCATTCATG 563  
3890 GTTCTCCAGAGAGCTGTGAATCGGCTCTGTGTGCTGAGAAAGCCCATTCATG 3949  
564 AGAGGCAAGGCCAGTGGG-TCCCCCAACTCCCCAGCCCCCTCTCCACAATGCACAGC 622  
3950 AGAGACGAGGTCCAGTGGGTCTCTCGTACTCCAGACCCCTCTCCACAATGCCCCC 4009

QY 623 CTCCCGCCCTCATCCCGCCCGCCCGCCCGCCCGCCCG 654  
Db 4010 TGTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 4041  
RESULT 12  
HUMCOL2A1Z  
LOCUS 31001 bp DNA linear PRI 03-AUG-1995  
DEFINITION Human pro-alpha1 type II collagen (COL2A1) gene exons 1-54,  
complete cds.  
ACCESSION U0347  
VERSION 1.0  
KEYWORDS alpha-1 type II collagen.  
SOURCE Homo sapiens  
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (sites)  
AUTHORS Cheah,K.S., Stoker,N.G., Griffin,J.R., Grosveld,F.G. and Solomon,E.  
TITLE Identification and characterization of the human type II collagen  
gene (COL2A1)  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 82 (9), 2555-2559 (1985)  
MEDLINE 85190534  
PUBMED 3857598  
REFERENCE 2 (sites)  
AUTHORS Baldwin,C.T., Reginato,A.M., Smith,C., Jimenez,S.A. and  
Prockop,D.J.  
TITLE Structure of cDNA clones coding for human type II procollagen. The  
alpha 1(II) chain is more similar to the alpha 1(I) chain than two  
other alpha chains of fibrillar collagens  
JOURNAL Biochem. J. 262 (2), 521-528 (1989)  
MEDLINE 90026318  
PUBMED 2803268  
REFERENCE 3 (sites)  
AUTHORS Vikkula,M. and Peltonen,L.  
TITLE Structural analyses of the polymorphic area in type II collagen  
gene  
JOURNAL FEBS Lett. 250 (2), 171-174 (1989)  
MEDLINE 89325561  
PUBMED 2753125  
REFERENCE 4 (sites)  
AUTHORS Ryan,M.C., Sieraski,M. and Sandell,L.J.  
TITLE The human type II procollagen gene: identification of an additional  
protein-coding domain and location of potential regulatory  
sequences in the promoter and first intron  
JOURNAL Genomics 8 (1), 41-48 (1990)  
MEDLINE 91184811  
PUBMED 2081599  
REFERENCE 5 (sites)  
AUTHORS Huang,M.C., Seyer,J.M., Thompson,J.P., Spinella,D.G., Cheah,K.S.  
and Kang,A.H.  
TITLE Genomic organization of the human procollagen alpha 1(II) collagen  
gene  
JOURNAL Eur. J. Biochem. 195 (3), 593-600 (1991)  
MEDLINE 91153296  
PUBMED 1999183  
REFERENCE 6 (sites)  
AUTHORS Vikkula,M., Metsaranta,M., Syvanen,A.C., Ala-Kokko,L., Vuorio,E.  
and Peltonen,L.  
TITLE Structural analysis of the regulatory elements of the type-II  
procollagen gene. Conservation of promoter and first intron  
sequences between human and mouse  
JOURNAL Biochem. J. 285 (Pt 1), 287-294 (1992)  
MEDLINE 92344585  
PUBMED 1637314  
REFERENCE 7 (sites)  
AUTHORS Ala-Kokko,L., Kvist,A.P., Metsaranta,M., Kivirikko,K.I., de  
Crombrughe,B., Prockop,D.J. and Vuorio,E.  
TITLE Conservation of the sizes of 53 introns and over 100 intronic  
sequences for the binding of common transcription factors in the  
human and mouse genes for type II procollagen (COL2A1)  
JOURNAL Biochem. J. 308 (Pt 3), 923-929 (1995)





2193 GTTCTCCGACAGACCTGTGAATCGGCTCTGTGTGCTCGAGAAAGCCCAATTCATG 2252  
564 AGAGCAAGGCCAGTGGG-TCCCCCAACTCCCGACCCCTCTCCCAATGCACAGC 622  
2253 AGAGCAGAGTCCAGTGGTCTCTGTATCTCCAGACCCCTCTCCCAATGCACAGC 2312  
623 CTCGCCGCTTCATCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 654  
2313 TGTGCGCGCGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 2344

SULT 13  
034871  
CUS  
FINITION Sequence 5 from Patent DE19909156.  
CESSION AX034871  
RSTON AX034871.1 GI:11190811  
YWORDS  
URCE  
ORGANISM  
FERENCE 1 (bases 1 to 282)  
AUTHORS Vornlocher,H.P., Bauer,B., Simandi,C., Achsel,T., Huels,C. and Luehrmann,R.  
JOURNAL Patent: DE 19909156-A 5 07-SEP-2000;  
AVENTIS RES & TECH GMBH & CO (DE)  
ATRES Location/Qualifiers  
source  
1..282  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="pra-mRNA"

SE COUNT 58 a 80 c 69 g 75 t  
IGIN  
Query Match 12.1%; Score 114.2; DB 6; Length 282;  
Best Local Similarity 87.4%; Pred.No.3.7e-16;  
Matches 125; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
727 AGCTCTAGTGGATCCCGGCTGCAGATCTGTAGGCGGAGTAGTCCAGGGTTCTCTTG 786  
139 AGCACTTCGACGCCAAGCTGTGCAGCTTAGGCGGAGTAGTCCAGGGTTCTCTTG 198  
787 ATGATGCATATCTATCTGCTCCCTTTTTCACAGCTCGCGGTTGAGGACAAACTCT 846  
199 ATGATGCATATCTATCTGCTCCCTTTTTCACAGCTCGCGGTTGAGGACAAACTCT 258  
847 TCGCGCTCTTCCAGTGGGATC 869  
259 TCGCGGCTCTTCCAGTGGGATC 281

SULT 14  
MCOLARAB  
CUS HUMCOLAAB 758 bp DNA linear PRI 01-NOV-1994  
FINITION Human alpha-1 collagen type II gene, exon 1, clone HC-2-C.  
CESSION M25698  
RSTON M25698.1 GI:180872  
YWORDS collagen.  
URCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Nunez,A.M., Kohn,K., Martin,G.R. and Yamada,Y.  
1 (bases 1 to 758)  
Promoter region of the human pro-alpha 1(II)-collagen gene  
Gene 44 (1), 11-16 (1986)  
MEDLINE 87031574  
PUBMED 3021582  
Location/Qualifiers  
source  
1..758  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="12q13"

gene 403..758  
/gene="COL2A1"  
exon 403..642  
/gene="COL2A1"  
/note="G00-119-063"  
/number=1  
CDS 558..>642  
/gene="COL2A1"  
/note="precursor"  
/codon\_start=1  
/product="alpha-1 type II collagen"  
/protein\_id="AA52051.1"  
/db\_xref="GI:553237"  
/db\_xref="GDB:G00-119-063"  
/translation="MIRLQAPQTLVLLVLAAYLRCQGDV"  
558..632  
/gene="COL2A1"  
/note="G00-119-063"  
633..>642  
/gene="COL2A1"  
/product="alpha-1 type II collagen"  
/note="G00-119-063"  
643..>758  
/gene="COL2A1"  
/note="G00-119-063"  
/number=1  
BASE COUNT 86 a 277 c 272 g 123 t  
ORIGIN Chromosome 12q14.3.  
Query Match 12.1%; Score 113.8; DB 9; Length 758;  
Best Local Similarity 87.9%; Pred.No.3.7e-16;  
Matches 124; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 2 GCCGGCTCGGGTGAGCCCTGATCCGCTCGGGGCTCCCGAGTCGCTGCTGCTGAC 61  
Db 539 CGAGCCCCGGGTGAGCCATGATTCGCTCGGGGCTCCCGAGTCGCTGCTGAC 598  
QY 62 GCTGCTCATCGCGGCTCTACGCTGTGAGGCGCAGGATGCCGTAAGTCGCCCGCC 121  
Db 599 GCTGCTGTGCGGCTGCTCTCGGTGTGAGGCGCAGGATGTCGGTAAGTCCTCCCGCC 658  
QY 122 CCTGCTACTTCCCTGACTT 142  
Db 659 CCTGCTGCTGCTGCTT 679

RESULT 15  
AX034878  
LOCUS AX034878 244 bp DNA linear PAT 15-NOV-2000  
DEFINITION Sequence 12 from Patent DE19909156.  
ACCESSION AX034878  
VERSION AX034878.1 GI:11190818  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 244)  
AUTHORS Vornlocher,H.P., Bauer,B., Simandi,C., Achsel,T., Huels,C. and Luehrmann,R.  
JOURNAL Patent: DE 19909156-A 12 07-SEP-2000;  
AVENTIS RES & TECH GMBH & CO (DE)  
FEATURES Location/Qualifiers  
source  
1..244  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="pra-mRNA"

BASE COUNT 47 a 68 c 62 g 67 t  
ORIGIN  
Query Match 12.0%; Score 113; DB 6; Length 244;  
Best Local Similarity 95.9%; Pred.No.6.4e-16;  
Matches 116; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

NY 757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTTATCCTGTCCCTTTTTT 816  
b 85 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTTATCCTGTCCCTTTTTT 144  
NY 817 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGGGTCTTCCAGTGGGGATCGACGGTA 876  
b 145 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGGGTCTTCCAGTGGGGATCCAAAGTGA 204  
NY 877 T 877  
b 205 T 205

earch completed: February 19, 2003, 23:17:07  
ob time : 3357.96 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

(nucleic - nucleic search, using sw model)

on: February 19, 2003, 19:26:38 ; Search time 573.739 Seconds  
(without alignments)  
3705.323 Million cell updates/sec

file: US-09-808-388-7  
fect score: 944  
quence: 1 tgcggcctcggtgagcc.....cgtgctgactgggaacc 944

oring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

arched: 2185239 seqs, 112599159 residues

tal number of hits satisfying chosen parameters: 4370478

num DB seq length: 0  
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

tabase : N Geneseq 101002.\*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	944	100.0	944	22	Sequence conferrin
2	216.6	22.9	7175	22	Vector pRK57SA-bet
3	171.8	18.2	372	24	Matrix metallopro
4	169.6	18.0	3479	20	Sprague Dawley rat
5	169.6	18.0	5276	20	CPE-TfA transgene
6	169.6	18.0	7664	20	CPE-lacZ construct
7	144	15.3	8083	23	Mouse fibroblast g
8	114.2	12.1	246	21	Splice variant seq
9	114.2	12.1	246	21	Splice variant DNA

10	114.2	12.1	282	21	AAA98324	Splice variant SEQ
11	113	12.0	244	21	AAA98335	Splice variant SEQ
12	113	12.0	244	21	AAA98341	Splice variant DNA
13	113	12.0	283	21	AAA98322	Splice variant SEQ
14	112.8	11.9	246	21	AAA98336	Splice variant SEQ
15	112.8	11.9	246	21	AAA98342	Splice variant DNA
16	112.8	11.9	283	21	AAA98323	Splice variant SEQ
17	112.6	11.9	241	21	AAA98351	Splice variant DNA
18	112.6	11.9	277	21	AAA98333	Splice variant DNA
19	112.6	11.9	277	21	AAA98358	Minx pre-mRNA DNA
20	111.4	11.8	247	21	AAA98338	Splice variant SEQ
21	111.4	11.8	247	21	AAA98344	Splice variant DNA
22	111.4	11.8	256	21	AAA98339	Splice variant SEQ
23	111.4	11.8	256	21	AAA98340	Splice variant DNA
24	111.4	11.8	256	21	AAA98345	Splice variant DNA
25	111.4	11.8	256	21	AAA98346	Splice variant DNA
26	111.4	11.8	282	21	AAA98325	Splice variant SEQ
27	111.4	11.8	282	21	AAA98326	Splice variant SEQ
28	111.4	11.8	282	21	AAA98327	Splice variant SEQ
29	109.2	11.6	8062	22	AA04946	Plasmid PRK76 used
30	109.2	11.6	8153	22	AA04945	Adenovirus type 2
31	104.4	11.1	120	22	AA04915	Nucleotide sequenc
32	104.4	11.1	120	22	AA04915	Nucleotide sequenc
33	104.4	11.1	1240	21	AA05065	Adenovirus 5 tripa
34	104.4	11.1	1240	24	ABA94272	Nucleotide sequenc
35	104.4	11.1	7231	21	AAA59090	Nucleotide sequenc
36	104.4	11.1	7231	24	ABA94286	Nucleotide sequenc
37	104.4	11.1	7960	21	AAA59072	Nucleotide sequenc
38	104.4	11.1	7960	24	ABA94274	Nucleotide sequenc
39	104.4	11.1	7989	21	AAA59075	Nucleotide sequenc
40	104.4	11.1	7989	24	ABA94277	Nucleotide sequenc
41	104.4	11.1	8383	21	AAA59071	Nucleotide sequenc
42	104.4	11.1	8383	24	ABA94273	Nucleotide sequenc
43	104.4	11.1	8484	21	AAA59091	Nucleotide sequenc
44	104.4	11.1	10332	20	AAV33921	Nucleotide sequenc
45	104.4	11.1	30385	24	ABK49011	Adenoviral vector

## ALIGNMENTS

RESULT 1  
AA164309  
ID AA164309 standard; DNA; 944 BP.  
XX  
AC AA164309;  
XX  
DT 15-NOV-2001 (first entry)  
XX  
DE Sequence conferring chondrocyte-specific gene expression.  
XX  
KW PPAR response element; antiinflammatory; antiarthritic; cytostatic;  
KW cardiant; nootropic; promoter; arthritis; tumour; PLA2IIA;  
KW peroxisome proliferator activated receptor; chondrocyte-specific;  
KW secreted non-pancreatic phospholipase A2; ds.  
XX  
OS Unidentified.  
XX  
PN WO200168845-A2.  
XX  
PD 20-SEP-2001.  
XX  
PP 14-MAR-2001; 2001WO-FR00759.  
XX  
PR 14-MAR-2000; 2000FR-0003262.  
PR 13-APR-2000; 2000US-0196959.  
PA (AVET ) AVENTIS PHARMA SA.  
XX  
PI Massad C, Berenbaum F, Olivier J, Salvat C, Bereziat G;  
DR WPI; 2001-582451/65.  
XX

PT New hybrid promoter induced by inflammation, useful in gene therapy of  
PT arthritis, comprises peroxisome proliferator activated receptor  
XX response element and promoter of secreted phospholipase A2 -  
PS Claim 12; Page 52; 52pp; French.  
XX  
CC The present invention relates to a hybrid promoter comprising (i) a PPAR  
CC (peroxisome proliferator activated receptor) response element (SPRE); and  
CC (ii) at least part of the promoter of the PLAZSIIA (secreted  
CC non-pancreatic phospholipase A2) gene. The hybrid promoter can be used to  
CC regulate expression of therapeutic transgenes, for experimental,  
CC clinical, therapeutic or diagnostic use, especially in chondrocytes for  
CC treatment of arthritis, but also in bone, muscle, liver, heart, the  
CC nervous system and tumours. The present sequence is a sequence which  
CC confers chondrocyte-specific sequence expression, which was used in the  
CC present invention.  
XX  
XQ Sequence 944 BP; 165 A; 335 C; 232 G; 212 T; 0 other;  
Query Match 100.0%; Score 944; DB 22; Length 944;  
Best Local Similarity 100.0%; Pred. No. 1.3e-223;  
Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
2y 1 TGCCGGCTCGCGGTGAGCCCTGATCCGCTCGGGGTCCCAAGTCTGCTGCTCTGA 60  
2b 1 TGCCGGCTCGCGGTGAGCCCTGATCCGCTCGGGGTCCCAAGTCTGCTGCTCTGA 60  
2y 61 CGCTCTCATCGCGCGTCTTAGGTGTGAGGCGCAGGATGCCGTAAGTCCCGCGG 120  
2b 61 CGCTCTCATCGCGCGTCTTAGGTGTGAGGCGCAGGATGCCGTAAGTCCCGCGG 120  
2y 121 CCCCTGCTACTTCCCTGACTGTGACCCCTTCTCTCTACTCTCCCTCCCAAGTACTAG 180  
2b 121 CCCCTGCTACTTCCCTGACTGTGACCCCTTCTCTCTACTCTCCCTCCCAAGTACTAG 180  
2y 181 GATCCCCCTAGAGCTGAGATCTGGGATGGGAGGATGGCTTCAGATGGGCTGAAC 240  
2b 181 GATCCCCCTAGAGCTGAGATCTGGGATGGGAGGATGGCTTCAGATGGGCTGAAC 240  
2y 241 CCGTCCCGTATTATTAACTGTTCTCTGAGAGCTGTGAATCGGGCTCTGTATGC 300  
2b 241 CCGTCCCGTATTATTAACTGTTCTCTGAGAGCTGTGAATCGGGCTCTGTATGC 300  
2y 301 GCTTGAGAAAGCCCAATCATGAGAGCAAGGCCAGTGGTCCCGCAACTCCCGGACC 360  
2b 301 GCTTGAGAAAGCCCAATTCATGAGAGCAAGGCCAGTGGTCCCGCAACTCCCGGACC 360  
2y 361 CCCCTCTCCCAATGACAGCTCCCGGCTCATCCCGGCTCATCCCGGCTCCCGGCT 420  
2b 361 CCCCTCTCCCAATGACAGCTCCCGGCTCATCCCGGCTCATCCCGGCTCCCGGCT 420  
2y 421 CTGCGGCCACCTTCAGATCGATCTGGGATGGCAGCGATGGCTTCAGATGGGCTGAAC 480  
2b 421 CTGCGGCCACCTTCAGATCGATCTGGGATGGCAGCGATGGCTTCAGATGGGCTGAAC 480  
2y 481 CCGTCCCGTATTATTAACTGTTCTCTGAGAGCTGTGAATCGGGCTCTGTATGC 540  
2b 481 CCGTCCCGTATTATTAACTGTTCTCTGAGAGCTGTGAATCGGGCTCTGTATGC 540  
2y 541 GCTTGAGAAAGCCCAATTCATGAGAGCAAGGCCAGTGGTCCCGCAACTCCCGGACC 600  
2b 541 GCTTGAGAAAGCCCAATTCATGAGAGCAAGGCCAGTGGTCCCGCAACTCCCGGACC 600  
2y 601 CCCCTCTCCCAATGACAGCTCCCGGCTCATCCCGGCTCATCCCGGCTCCCGGCT 660  
2b 601 CCCCTCTCCCAATGACAGCTCCCGGCTCATCCCGGCTCATCCCGGCTCCCGGCT 660  
2y 661 CCGTCCCGGCTCATCCAGATCTCCAGCTAGAGGATCTCGGACTCTAGGGTTCGAATCGAT 720  
2b 661 CCGTCCCGGCTCATCCAGATCTCCAGCTAGAGGATCTCGGACTCTAGGGTTCGAATCGAT 720  
2y 721 AAGCCAGCTCTAGTGGATCCCGGGCTGAGATCTGTAGGGCGCAGTAGTCCAGGGTT 780  
2b 721 AAGCCAGCTCTAGTGGATCCCGGGCTGAGATCTGTAGGGCGCAGTAGTCCAGGGTT 780

Db 721 AAGCCAGCTCTAGTGGATCCCGGGCTGAGATCTGTAGGGCGCAGTAGTCCAGGGTT 780  
Qy 781 TCCTTGATGATGTCATACATTATCCTGTCTCCCTTTTTCACAGCTCGCGGTTGAGGACA 840  
Db 781 TCCTTGATGATGTCATACATTATCCTGTCTCCCTTTTTCACAGCTCGCGGTTGAGGACA 840  
Qy 841 AACTTCTCGGGTCTTTCAGTGGGATCGAGGTATCGATTAAGCTTGATCTGTGAC 900  
Db 841 AACTTCTCGGGTCTTTCAGTGGGATCGAGGTATCGATTAAGCTTGATCTGTGAC 900  
Qy 901 ATGGCGGATCCCGTCTGTTTACAACTGCTGACTGGGAAACCC 944  
Db 901 ATGGCGGATCCCGTCTGTTTACAACTGCTGACTGGGAAACCC 944  
RESULT 2  
AAD04927  
ID AAD04927 standard; DNA; 7175 BP.  
XX  
AC AAD04927;  
XX  
DT 17-JUL-2001 (first entry)  
XX  
DE Vector pRK57SA-beta used as a gene trap in embryonic stem (ES) cells.  
XX  
KW Gene trapping construct; conditional mutation; unidirectional inversion;  
KW recombinase recognition sequence; RRS; disruption cassette;  
KW selection cassette; transgenic organism; vector pRK57SA-beta;  
KW embryonic stem; ES; splice acceptor; SA; fusion gene; beta-galactosidase;  
KW neomycin-phosphotransferase; bovine polyadenylation signal; bpa;  
KW mutant loxP site; ds.  
XX  
OS Chimeric - Escherichia coli.  
OS Chimeric - Adenovirus.  
OS Chimeric - Bos sp.  
OS Chimeric - Bacteriophage P1.  
XX  
PH Key Location/Qualifiers  
FT misc\_feature 446..479  
FT /tag= a  
FT /notes "Bacteriophage P1 loxP mutant site (lox66)"  
FT  
FT misc\_feature 525..628  
FT /tag= b  
FT /notes "Adenovirus type 2 splice acceptor from  
FT exon 2 of the major late region"  
FT  
FT misc\_feature 667..4548  
FT /tag= c  
FT /note= "E. coli beta-galactosidase and  
FT neomycin-phosphotransferase fusion gene"  
FT  
FT misc\_feature 4561..4843  
FT /tag= d  
FT /notes "Bovine growth hormone polyadenylation signal"  
FT  
FT misc\_feature 4869..4902  
FT /tag= e  
FT /note= "Bacteriophage P1 loxP mutant site (lox71)"  
XX  
WO200129208-A1.  
PN  
XX  
PD 26-APR-2001.  
XX  
PF 16-OCT-2000; 2000WO-BP10162.  
XX  
XX 16-OCT-1999; 99EP-0120592.  
PR 27-OCT-1999; 99US-0162016.  
XX  
XX (ARTE-) ARTEMIS PHARM GMBH.  
PA (FRAN-) FRANKEN BIOTECHNOLOGIE AG.  
XX  
XX Kuehn R, Von Melchener H, Altschmied J;  
PI  
XX WPI; 2001-308486/32.  
DR  
XX New gene trapping construct capable of causing conditional mutations in  
PT

genes, comprises functional DNA segment inserted in sense or antisense direction relative to gene to be trapped -  
Example 1; Page 53-55; 78pp; English.

The present invention relates to a conditional gene trapping construct capable of causing conditional mutations in genes. The gene trapping construct comprises two functional DNA segments, each being flanked by two recombinase recognition sequences (RRSs) specific to site specific recombinase which is capable of unidirectional inversion of double standard DNA segment. One of the DNA segment (disruption cassette) is inserted in antisense orientation relative to the transcriptional orientation of the gene to be trapped. The other DNA segment (selection cassette) is inserted in sense direction relative to the transcriptional orientation of the gene to be trapped. The cell comprising the gene trapping construct is useful for the identification and/or isolation of genes. The transgenic organism comprising the gene trapping construct is useful to study gene function at various developmental stages. The gene trapping construct is useful for mutationally inactivating all cellular genes. The present sequence is vector pRK578A-beta which is used as a gene trap in embryonic stem (ES) cells. The vector pRK578A-beta comprises adenovirus type 2 splice acceptor (SA) from exon 2 of the major late region, a fusion gene of *Escherichia coli* beta-galactosidase and neomycin-phosphotransferase (beta-geo), the transcription termination and bovine growth hormone polyadenylation signal (bphA). The SA-beta-geo cassette is flanked by mutant loxP sites (5' lox66 and 3' lox71) in opposite orientation.

Sequence 7175 BP; 1614 A; 1892 C; 2002 G; 1667 T; 0 other;

Query Match 22.9%; Score 216.6; DB 22; Length 7175;  
Best Local Similarity 99.2%; Pred. No. 1e-43;  
Matches 219; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
772 AGCCAGCTCTAGTGGATCCCGGGCTGCGAGATCTGTAGGGCGAGTATCCAGGGTTT 781  
488 AGGCTAGAACTAGTGGATCCCGGGCTGCGAGATCTGTAGGGCGAGTATCCAGGGTTT 547  
792 CTTGATGATGTCATCTTACATCTTCCCTTTTTCACAGCTCGGGTTCAGGACAA 841  
548 CTTGATGATGTCATCTTACATCTTCCCTTTTTCACAGCTCGGGTTCAGGACAA 607  
842 ACTCTTCGGGTCTTTCAGTGGGATCGACGGTATCGATAAGCTTGATGATCTGTGACA 901  
608 ACTCTTCGGGTCTTTCAGTGGGATCGACGGTATCGATAAGCTTGATGATCTGTGACA 667  
902 TGGCGGATCCCGTCTGTTTACACGCTGTGACTGGGAAACCC 944  
668 TGGCGGATCCCGTCTGTTTACACGCTGTGACTGGGAAACCC 710

RESULT 3  
K47530  
ABK47530 standard; DNA; 372 BP.  
ABK47530;  
18-JUN-2002 (first entry)  
Matrix metalloproteinase 19 (MMP-19) associated enhancer.

Transgenic animal; matrix metalloproteinase 19; MMP-19;  
extracellular matrix disorder; chondrogenic failure;  
osteogenic failure; osteoporosis; arthritis; synovitis; eye disease;  
malignant tumour; joint disease; bone disease; bone deformation;  
limb shortening; cranial deformation; defective bite; tooth elongation;  
enhancer; ds.

Rattus sp.  
WO200211530-A1.  
14-FEB-2002.

XX 08-AUG-2001; 2001WO-JP06826.  
XX 09-AUG-2000; 2000JP-0241748.  
XX (TAKA) TAKEDA CHEM IND LTD.  
XX Yoshimura K, Nishimura A, Nishida M, Hosono K;  
XX WPI; 2002-227106/28.  
XX Transgenic mammal containing foreign MMP-19 gene for use as a model for  
XX bone and cartilage diseases -  
XX Example 1; Page 36; 46pp; Japanese.  
XX The invention describes a non-human transgenic mammal containing a  
XX recombinant DNA encoding a foreign matrix metalloproteinase 19 (MMP-19)  
XX gene or its modified form. Identification of agents for the treatment  
XX and prevention of extracellular matrix disorders including chondrogenic  
XX failure, osteogenic failure, osteoporosis, arthritis deformans,  
XX rheumatoid arthritis, synovitis, metabolic arthritis, eye disease,  
XX malignant tumours, and associated complications. The transgenic mammals  
XX are a model for joint and bone diseases including deformation and  
XX shortening of limbs, cranial deformation, defective bite, tooth  
XX elongation, and defects of lumbar and tail vertebrae. This sequence  
XX represents an enhancer associated with the creation of the transgenic  
XX animal expressing the recombinant MMP-19 protein.

Sequence 372 BP; 66 A; 123 C; 97 G; 86 T; 0 other;

Query Match 18.2%; Score 171.8; DB 24; Length 372;  
Best Local Similarity 82.9%; Pred. No. 5.8e-33;  
Matches 208; Conservative 0; Mismatches 42; Indels 1; Gaps 1;  
QY 410 CCGCTGCGCGCTGCGCGCACCTTCAGATGATCTGGGATGCGAGGATGGCTTCCAGA 469  
Db 103 CTTGAGGCTTGTGGGCTTGTGGGATGCGAGGATGGCTTCCAGA 162  
QY 470 TGGGCTGAACCTGCGCGCTTATTTAACTGTTCTCGTGGAGAGCTGTGATCGG 529  
Db 163 TGGGCTGAACCTGCGCGCTTATTTAACTGTTCTCGTGGAGAGCTGTGATCGG 222  
QY 530 GCTCTGTATGCGCTTGAGAAAAGCCCATTCATGAGAGCAAGGCCAGTGGGT-CCCC 588  
Db 223 GCTCTGTATGCGCTTGAGAAAAGCCCATTCATGAGAGCAAGGCCAGTGGGTCCCC 282  
QY 589 AACTCCCGAGCCCCCTTCCGCAATGACAGCTTCCCGGCCCTCATCCCCCCCCCAC 648  
Db 283 GACTCCCGAGCCCCCTTCCGCAATGACAGCTTCCCGGCCCTCATCCCCCCCCCAC 342  
QY 649 CCCCCGTGCCC 659  
Db 343 CTCCCGGGCTC 353

RESULT 4  
AAX90515  
ID AAX90515 standard; DNA; 3479 BP.  
XX AAX90515;  
XX 04-OCT-1999 (first entry)  
XX Sprague Dawley rat type II collagen promoter.  
XX Human; matrix metalloproteinase 13; MMP-13; collagenase 3; cartilage;  
XX degenerative disease; matrix degrading enzyme; MDE; transgenic mammal;  
XX promoter; ss.  
XX Rattus sp.  
XX WO9931969-A2.

X D 01-JUL-1999.  
X F 18-DEC-1998; 98WO-US27056.  
X R 19-DEC-1997; 97US-0994689.  
X R 19-DEC-1997; 97US-0068312.  
X A (AMHP ) AMERICAN HOME PROD CORP.  
X I Killar LM, Neuhold LA;  
X R WPI; 1999-468690/39.  
X T Transgenic mammals that express recombinant matrix-degrading  
T enzymes, used to study phenotypic changes associated with  
T cartilage-degenerative disease  
S Disclosure; Page 61-62; 70pp; English.  
X C The present invention describes transgenic mammals that express  
C recombinant matrix-degrading enzymes (MDE) in a temporally and spatially  
C regulated manner. The transgenic mammals produce phenotypic changes  
C associated with cartilage-degenerative disease if maintained under  
C conditions in which the recombinant gene is selectively expressed in  
C joint tissue. The transgenic animal models are useful for determining  
C the potential of a composition to counteract cartilage-degenerative  
C disease. The present sequence represents a promoter which can be  
C used in a transgenic mammal of the present invention.  
X Q Sequence 3479 BP; 704 A; 979 C; 985 G; 811 T; 0 other;  
Query Match 18.0%; Score 169.6; DB 20; Length 3479;  
Best Local Similarity 80.8%; Pred. No. 3.5e-32;  
Matches 210; Conservative; 0; Mismatches 49; Indels 1; Gaps 1;  
Y 410 CCCGTGCGCCGCTGCGCCACCTTCAGATCGATCTGGGATTGGCAGCGATGCTTCCAGA 469  
b 987 CCTCAGTCTCTCTTTGTGAGGCTTTTGTGCTTGGAGATTGGCAGCGATGCTTCCAGA 1046  
Y 470 TGGGCTGAACCCCTGCGCGTATTTATTTAACTGGTTCCTCTGAGAGCTGTGAATCGG 529  
b 1047 TGGGCTGAACCCCTGCGCGTATTTATTTAACTGGTTCCTCTGAGAGCTGTGAATCGG 1106  
Y 530 GCTCTGTATCGCTTCGAGAAAGCCCATTCATGAGAGCAAGGCCAGTGGGT-CCCCC 588  
b 1107 GCTCTGTATCGCTTCGAGAAAGCCCATTCATGAGAGCAAGGCCAGTGGGT-CCCCC 1166  
Y 589 AACTTCCCGACCCCGCTCTCCCAATATGACAGCTTCCCGCCCTCATCCCCCCCCCACC 648  
b 1167 GACTTCCCGACCCCGCTCTCCCAATATATATATATATATATATATATATATATATAT 1226  
Y 649 CCCCCGTGCGCGCTGCGCGC 668  
b 1227 CTCCCGGGCTCGGCGCGCGC 1246  
ESULT 5  
AX30502  
D AX90502 standard; DNA; 5276 BP.  
X C AX90502;  
X T 04-OCT-1999 (first entry)  
X E CPE-tTA transgene.  
X W Human; matrix metalloproteinase 13; MMP-13; collagenase 3; cartilage;  
W degenerative disease; matrix degrading enzyme; MDE; transgenic mammal;  
W ss.  
X S Synthetic.  
S Rattus sp.

XX WO9931369-A2.  
XX 01-JUL-1999.  
XX 18-DEC-1998; 98WO-US27056.  
XX 19-DEC-1997; 97US-0994689.  
XX 19-DEC-1997; 97US-0068312.  
XX (AMHP ) AMERICAN HOME PROD CORP.  
XX Killar LM, Neuhold LA;  
XX WPI; 1999-468690/39.  
XX Transgenic mammals that express recombinant matrix-degrading  
PT enzymes, used to study phenotypic changes associated with  
PT cartilage-degenerative disease  
XX Example 3; Page 64-66; 70pp; English.  
XX The present invention describes transgenic mammals that express  
CC recombinant matrix-degrading enzymes (MDE) in a temporally and spatially  
CC regulated manner. The transgenic mammals produce phenotypic changes  
CC associated with cartilage-degenerative disease if maintained under  
CC conditions in which the recombinant gene is selectively expressed in  
CC joint tissue. The transgenic animal models are useful for determining  
CC the potential of a composition to counteract cartilage-degenerative  
CC disease. The present sequence represents a transgene with rat type II  
CC collagen promoter driving expression of the tetracycline repressor -  
CC VP16 activator fusion protein (TA), followed by an SV40 splice and  
CC polyadenylation signal, where the transgene is designated CPE-tTA.  
XX Sequence 5276 BP; 1219 A; 1365 C; 1398 G; 1294 T; 0 other;  
SQ Query Match 18.0%; Score 169.6; DB 20; Length 5276;  
Best Local Similarity 80.8%; Pred. No. 3.9e-32;  
Matches 210; Conservative; 0; Mismatches 49; Indels 1; Gaps 1;  
QY 410 CCCGTGCGCCGCTGCGCGACCTTCAGATCGATCTGGGATTGGCAGCGATGCTTCCAGA 469  
Db 988 CTTCACTCTCTCTTTGTGAGGCTTTTGTGAGGATTGGCAGCGATGCTTCCAGA 1047  
QY 470 TGGGCTGAACCCCTGCGCGTATTTATTTAACTGGTTCCTCTGAGAGCTGTGAATCGG 529  
Db 1048 TGGGCTGAACCCCTGCGCGTATTTATTTAACTGGTTCCTCTGAGAGCTGTGAATCGG 1107  
QY 530 GCTCTGTATCGCTTCGAGAAAGCCCATTCATGAGAGCAAGGCCAGTGGGT-CCCCC 588  
Db 1108 GCTCTGTATCGCTTCGAGAAAGCCCATTCATGAGAGCAAGGCCAGTGGGT-CCCCC 1167  
QY 589 AACTTCCCGACCCCGCTCTCCCAATATGACAGCTTCCCGCCCTCATCCCCCCCCCACC 648  
Db 1168 GACTTCCCGACCCCGCTCTCCCAATATATATATATATATATATATATATATATATAT 1227  
QY 649 CCCCCGTGCGCGCTGCGCGC 668  
Db 1228 CTCCCGGGCTCGGCGCGCGC 1247  
RESULT 6  
AX90503  
ID AX90503 standard; DNA; 7664 BP.  
XX AX90503;  
XX AC AX90503;  
XX CPE-tTA transgene.  
XX 04-OCT-1999 (first entry)  
XX CPE-lacZ construct.  
XX Human; matrix metalloproteinase 13; MMP-13; collagenase 3; cartilage;  
KW degenerative disease; matrix degrading enzyme; MDE; transgenic mammal;

ss.

Synthetic.  
Rattus sp.

WO9931969-A2.

01-JUL-1999.

18-DEC-1998; 98WO-US27056.

19-DEC-1997; 97US-0994689.

19-DEC-1997; 97US-0068312.

(AMHP) AMERICAN HOME PROD CORP.

Killar LM, Neuhold LA;

WPI; 1999-468690/39.

Transgenic mammals that express recombinant matrix-degrading enzymes, used to study phenotypic changes associated with cartilage-degenerative disease

Example 3; Page 66-68; 70pp; English.

The present invention describes transgenic mammals that express recombinant matrix-degrading enzymes (MDE) in a temporally and spatially regulated manner. The transgenic mammals produce phenotypic changes associated with cartilage-degenerative disease if maintained under conditions in which the recombinant gene is selectively expressed in joint tissue. The transgenic animal models are useful for determining the potential of a composition to counteract cartilage-degenerative disease. The present sequence represents a CPB-lacZ construct with the rat type II collagen promoter driving expression of the beta-galactosidase (lacZ) gene followed by a beta-globin splice and polyadenylation signal.

Sequence 7664 BP; 1654 A; 2045 C; 2089 G; 1876 T; 0 other;

Query Match 18.0%; Score 169.6; DB 20; Length 7664;

Best Local Similarity 80.8%; Pred. No. 4.3e-32;

Matches 210; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

410 CCCGTCGCCGCTGCGCCACCTTCAGATCGATCTGGGATTGGCAGCGATGGCTCCAGA 469

988 CTTCACTCTCTTTGTGAGGCTTGTTCGTTGAGGATTGGCAGCGATGGCTCCAGA 1047

470 TGGGCTGAACCCCTGCCGATTTATTAACTGGTCTCTCGTGAGAGCTGTGAATCGG 529

1048 TGGGCTGAACCCCTGCCGATTTATTAACTGGTCTCTCGTGAGAGCTGTGAATCGG 1107

530 GCTCTGTATGGCTTGAGAAAGCCCATTCATGAGAGGCAAGCCCGAGTGGGT-CCCC 588

1108 GCTCTGTATGGCTTGAGAAAGCCCATTCATGAGAGGCAAGCCCGAGTGGGTCCCC 1167

589 AACTCCCGACCCCTCTCCACATGACAGCTCCCGCCCTCATCCCCCCCCCACC 648

1168 GACTCCCGACCCCTCTCCACATGATATCCCTCTCTGTCGCGCTGCGCCAC 1227

649 CCGCGTCCGCTGCGCG 668

1228 CTCGCGGCTCGGCGCCGC 1247

RESULT 7

UD34802

AAD34802 standard; cDNA; 8083 BP.

AAD34802;

16-JUL-2002 (first entry)

Mouse fibroblast growth factor receptor 3 (FGFR3) mutant cDNA.

Mouse; chondrodysplasia; achondroplasia; transgenic mouse; therapy; fibroblast growth factor receptor 3; FGFR3; limb; midface hypoplasia; large skull; drug screening; drug development; transgenic; mutant; gene; ss.

Mus sp.

Synthetic.

Key Location/Qualifiers

CD5 5018..7423

FT /\*tag= a

FT /product= "Mouse FGFR3 mutant protein"

XX US6265632-B1.

XX 24-JUL-2001.

XX 26-AUG-1999; 99US-0383630.

XX 27-AUG-1998; 98IL-0125958.

XX (YEDA) YEDA RES & DEV CO LTD.

XX (PROC-) PROCHON BIOTECH LTD.

XX Yayon A, Segev O;

XX WPI; 2001-463946/50.

XX P-PSDB; AAE21977.

New transgenic mice having a genetically modified fibroblast growth factor receptor gene, useful as a model for human chondrodysplasia, e.g. achondroplasia characterized by shortening of the limbs, midface hypoplasia or large skull -

Example; Column 45-54; 49pp; English.

The invention relates to an animal model for chondrodysplasia, more particularly, to a transgenic mouse model for achondroplasia. This transgenic mouse contains a fibroblast growth factor receptor 3 (FGFR3) gene including a G to A point mutation changing Gly to Arg in codon 380 in its genome. The transgenic mouse is useful as a model for FGFR-associated chondrodysplasia, particularly FGFR3 achondroplasia, e.g. shortening of the limbs, midface hypoplasia and large skull. This model may be exploited to gain better understanding of the disease and as an experimental model with which experimental therapy to chondrodysplasias can be exercised. The transgenic mouse is particularly useful as a tool for screening, developing and evaluating drugs with a potential of relieving or abolishing chondrodysplasia syndromes and/or symptoms. The present sequence is mouse FGFR3 mutant cDNA.

Sequence 8083 BP; 1680 A; 2186 C; 2533 G; 1684 T; 0 other;

Query Match 15.3%;

Best Local Similarity 96.7%; Pred. No. 9.2e-26;

Matches 147; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 732 TAGTGGATCCCGGGCTGCAGATCTGTAGGGCGCAGTAGTCCAGGGTTCTTGATGAT 791

Db 4867 TAGTGGATCCCGGGCTGCAGATCCCGGGCGCAGTAGTCCAGGGTTCTTGATGAT 4926

QY 792 GTCATACTTATCTGTCCTTTTTCACAGCTCGCGGTTGAGGACAACTCTTCGGG 851

Db 4927 GTCATACTTATCTGTCCTTTTTCACAGCTCGCGGTTGAGGACAACTCTTCGGG 4986

QY 852 GTCCTTCCAGTGGGGATCGACGGTATCGATAA 883

Db 4987 GTCCTTCCAGTGGGGATCGACGGTATCGATAA 5018

RESULT 8

AAA98337





DE19909156-A1.  
07-SEP-2000.  
02-MAR-1999; 99DE-1009156.  
02-MAR-1999; 99DE-1009156.  
(AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.  
Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;  
WPI; 2000-588345/56.  
Novel test system for detecting a splice reaction used to identify substances effective as fungicides, herbicides, pesticides and insecticides or to diagnose a disease -  
Disclosure; Page 12; 36pp; German.  
This invention describes a novel test system for detecting a splice reaction comprising at least 1 optionally similar immobilized nucleic acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel free detection system, at least a composition containing a splice component, a suitable detection probe, and if necessary other means of help. The method is used to identify substances, which are effective as fungicides, herbicides, pesticides and/or insecticides. The method can be used to diagnose cancer, a viral disease, Grave's disease, spinal muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene, hepatitis C infection and/or herpes simplex virus infection.  
Sequence 282 BP; 58 A; 80 C; 69 G; 75 T; 0 other;  
Query Match 12.1%; Score 114.2; DB 21; Length 282;  
Best Local Similarity 87.4%; Pred. No. 9.1e-19;  
Matches 125; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
727 AGCTCTAGTGGATCCCCGGGCTGCAGATCTGTAGGGCGCAGTAGTCCAGGGTTTCCTTG 786  
139 AGCATTTCTGCAGCCCAAGCTTGCTGCACGCTAGGGCGCAGTAGTCCAGGGTTTCCTTG 198  
787 ATGATGTCATATCTATCTGTCCTTTTTCACAGCTCGCGGTGAGGACAACTCT 846  
199 ATGATGTCATATCTATCTGTCCTTTTTCACAGCTCGCGGTGAGGACAACTCT 258  
847 TCGCGTCTTTCCAGTGGGATC 869  
259 TCGCGTCTTTCCAGTGGGATC 281  
SULT 11  
A98335  
AAA98335 standard; DNA; 244 BP.  
AAA98335;  
02-FEB-2001 (first entry)  
Splice variant SEQ ID NO: 16.  
RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer; pesticide; insecticide; diagnosis; viral disease; Grave's disease; spinal muscular atrophy; beta-thalassemia; DNA-RNA hybrid; ds.  
Unidentified.  
DE19909156-A1.  
07-SEP-2000.  
02-MAR-1999; 99DE-1009156.  
02-MAR-1999; 99DE-1009156.  
DE19909156-A1.  
07-SEP-2000.  
02-MAR-1999; 99DE-1009156.  
02-MAR-1999; 99DE-1009156.  
(AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.  
Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;  
WPI; 2000-588345/56.  
Novel test system for detecting a splice reaction used to identify substances effective as fungicides, herbicides, pesticides and insecticides or to diagnose a disease -  
Disclosure; Page 17; 36pp; German.  
This invention describes a novel test system for detecting a splice reaction comprising at least 1 optionally similar immobilized nucleic acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel free detection system, at least a composition containing a splice component, a suitable detection probe, and if necessary other means of help. The method is used to identify substances, which are effective as fungicides, herbicides, pesticides and/or insecticides. The method can be used to diagnose cancer, a viral disease, Grave's disease, spinal muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene, hepatitis C infection and/or herpes simplex virus infection.  
Sequence 244 BP; 47 A; 68 C; 62 G; 56 T; 11 U; 0 other;  
Query Match 12.0%; Score 113; DB 21; Length 244;  
Best Local Similarity 95.0%; Pred. No. 1.7e-18;  
Matches 115; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTATCTCCCTGCTTTT 816  
85 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTATCTCCCTGCTTTT 144  
817 TTCCACAGCTCGCGGTGAGGACAACTCTTCCGGTCTTTCAGTGGGATCGACGTA 876  
145 TTCCACAGCTCGCGGTGAGGACAACTCTTCCGGTCTTTCAGTGGGATCGACGTA 204  
877 T 877  
205 U 205  
RESULT 12  
AAA98341  
ID AAA98341 standard; DNA; 244 BP.  
XX  
AC AAA98341;  
XX  
DT 02-FEB-2001 (first entry)  
XX  
DE Splice variant DNA #1.  
XX  
KW RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer; pesticide; insecticide; diagnosis; viral disease; Grave's disease; spinal muscular atrophy; beta-thalassemia; ds.  
XX  
OS Unidentified.  
XX  
PN DE19909156-A1.  
XX  
PD 07-SEP-2000.  
XX  
PF 02-MAR-1999; 99DE-1009156.  
XX  
PR 02-MAR-1999; 99DE-1009156.  
XX  
PA (AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.  
XX  
PI Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;  
XX  
DR WPI; 2000-588345/56.  
XX

PT Novel test system for detecting a splice reaction used to identify  
PT substances effective as fungicides, herbicides, pesticides and  
PT insecticides or to diagnose a disease -  
XX  
XX Claim 19; Page 20; 36pp; German.  
XX  
XX This invention describes a novel test system for detecting a splice  
XX reaction comprising at least 1 optionally similar immobilized nucleic  
XX acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel  
XX free detection system, at least a composition containing a splice  
XX component, a suitable detection probe, and if necessary other means of  
XX help. The method is used to identify substances, which are effective as  
XX fungicides, herbicides, pesticides and/or insecticides. The method can  
XX be used to diagnose cancer, a viral disease, Grave's disease, spinal  
XX muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene,  
XX hepatitis C infection and/or herpes simplex virus infection.  
XX  
XX Sequence 244 BP; 47 A; 68 C; 62 G; 67 T; 0 other;

Query Match 12.0%; Score 113; DB 21; Length 244;  
Best Local Similarity 95.9%; Pred. No. 1.7e-18;  
Matches 116; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
XX  
XX 757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATCTCCCTTTT 816  
XX 85 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATCTCCCTTTT 144  
XX  
XX 817 TTCACAGCTCGCGTTGAGGACAACTCTTCGGGTCTTCCAGTGGGATCGACGGTA 876  
XX 145 TTCACAGCTCGCGTTGAGGACAACTCTTCGGGTCTTCCAGTGGGATCGACGGTA 204  
XX  
XX 877 T 877  
XX 205 T 205

RESULT 13  
AAA98322  
ID AAA98322 standard; DNA; 283 BP.  
XX  
XX AAA98322;  
XX  
XX 02-FEB-2001 (first entry)  
XX  
XX Splice variant SEQ ID NO: 3.  
XX  
XX RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer;  
XX pesticide; insecticide; diagnosis; viral disease; Grave's disease;  
XX spinal muscular atrophy; beta-thalassemia; ds.  
XX  
XX Unidentified.  
XX  
XX DE19909156-A1.  
XX  
XX 07-SEP-2000.  
XX  
XX 02-MAR-1999; 99DE-1009156.  
XX  
XX 02-MAR-1999; 99DE-1009156.  
XX  
XX (AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.  
XX  
XX Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;  
XX WPI; 2000-588345/56.  
XX  
XX Novel test system for detecting a splice reaction used to identify  
XX substances effective as fungicides herbicides, pesticides and  
XX insecticides or to diagnose a disease -  
XX  
XX Disclosure; Page 11; 36pp; German.  
XX  
XX This invention describes a novel test system for detecting a splice

CC reaction comprising at least 1 optionally similar immobilized nucleic  
CC acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel  
CC free detection system, at least a composition containing a splice  
CC component, a suitable detection probe, and if necessary other means of  
CC help. The method is used to identify substances, which are effective as  
CC fungicides, herbicides, pesticides and/or insecticides. The method can  
CC be used to diagnose cancer, a viral disease, Grave's disease, spinal  
CC muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene,  
CC hepatitis C infection and/or herpes simplex virus infection.  
XX  
XX Sequence 283 BP; 58 A; 80 C; 70 G; 75 T; 0 other;

Query Match 12.0%; Score 113; DB 21; Length 283;  
Best Local Similarity 95.9%; Pred. No. 1.8e-18;  
Matches 116; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
XX  
XX 757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATCTCCCTTTT 816  
XX 124 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATCTCCCTTTT 183  
XX  
XX 817 TTCACAGCTCGCGTTGAGGACAACTCTTCGGGTCTTCCAGTGGGATCGACGGTA 876  
XX 184 TTCACAGCTCGCGTTGAGGACAACTCTTCGGGTCTTCCAGTGGGATCGACGGTA 243  
XX  
XX 877 T 877  
XX 244 T 244

RESULT 14  
AAA98336  
ID AAA98336 standard; DNA; 246 BP.  
XX  
XX AAA98336;  
XX  
XX 02-FEB-2001 (first entry)  
XX  
XX Splice variant SEQ ID NO: 17.  
XX  
XX RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer;  
XX pesticide; insecticide; diagnosis; viral disease; Grave's disease;  
XX spinal muscular atrophy; beta-thalassemia; ds.  
XX  
XX Unidentified.  
XX  
XX DE19909156-A1.  
XX  
XX 07-SEP-2000.  
XX  
XX 02-MAR-1999; 99DE-1009156.  
XX  
XX 02-MAR-1999; 99DE-1009156.  
XX  
XX (AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.  
XX  
XX Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;  
XX WPI; 2000-588345/56.  
XX  
XX Novel test system for detecting a splice reaction used to identify  
XX substances effective as fungicides, herbicides, pesticides and  
XX insecticides or to diagnose a disease -  
XX  
XX Disclosure; Page 17; 36pp; German.

This invention describes a novel test system for detecting a splice  
reaction comprising at least 1 optionally similar immobilized nucleic  
acid with at least 1 nucleic acid (I) capable of splicing, at least 1 gel  
free detection system, at least a composition containing a splice  
component, a suitable detection probe, and if necessary other means of  
help. The method is used to identify substances, which are effective as  
fungicides, herbicides, pesticides and/or insecticides. The method can  
be used to diagnose cancer, a viral disease, Grave's disease, spinal

muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene,  
hepatitis C infection and/or herpes simplex virus infection.

Sequence 246 BP; 47 A; 67 C; 65 G; 67 T; 0 other;

Query Match 11.9%; Score 112.8; DB 21; Length 246;

Best Local Similarity 98.3%; Pred. No. 1.9e-18;

Matches 114; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

757 TCTAGGGCGCAGTAGTCCTGAGGTTTCCTTGATGATGTCATACCTATCTCTCCCTTTT 816

88 TCTAGGGCGCAGTAGTCCTGAGGTTTCCTTGATGATGTCATACCTATCTCTCCCTTTT 147

817 TTCCACAGCTCGCGTTGAGGACAACTCTTCGGGTCCTTCCAGTGGGATCGAC 872

148 TTCCACAGCTCGCGTTGAGGACAACTCTTCGGGTCCTTCCAGTGGGATCGGC 203

SULT 15

A98342

AAA98342 standard; DNA; 246 BP.

AAA98342;

02-FEB-2001 (first entry)

Splice variant DNA #2.

RNA aptamer; splice reaction; detection; fungicide; herbicide; cancer;  
pesticide; insecticide; diagnosis; viral disease; Grave's disease;  
spinal muscular atrophy; beta-thalassemia; ds.

Unidentified.

DE19909156-A1.

07-SEP-2000.

02-MAR-1999; 99DE-1009156.

02-MAR-1999; 99DE-1009156.

(AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.

Huels C, Bauer B, Simandi C, Luehrmann R, Achsel T, Vornlocher H;

WPI; 2000-588345/56.

Novel test system for detecting a splice reaction used to identify  
substances effective as fungicides, herbicides, pesticides and  
insecticides or to diagnose a disease -

Claim 19; Page 20; 36pp; German.

This invention describes a novel test system for detecting a splice  
reaction comprising at least 1 optionally similar immobilized nucleic  
acid with at least 1 nucleic acid (i) capable of splicing, at least 1 gel  
free detection system, at least a composition containing a splice  
component, a suitable detection probe, and if necessary other means of  
help. The method is used to identify substances, which are effective as  
fungicides, herbicides, pesticides and/or insecticides. The method can  
be used to diagnose cancer, a viral disease, Grave's disease, spinal  
muscular atrophy, beta-thalassemia, cancer caused by c-erb oncogene,  
hepatitis C infection and/or herpes simplex virus infection.

Sequence 246 BP; 47 A; 67 C; 65 G; 67 T; 0 other;

Query Match 11.9%; Score 112.8; DB 21; Length 246;

Best Local Similarity 98.3%; Pred. No. 1.9e-18;

Matches 114; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

757 TCTAGGGCGCAGTAGTCCTGAGGTTTCCTTGATGATGTCATACCTATCTCTCCCTTTT 816

Db 88 TCTAGGGCGCAGTAGTCCTGAGGTTTCCTTGATGATGTCATACCTATCTCTCCCTTTT 147

Qy 817 TTCCACAGCTCGCGTTGAGGACAACTCTTCGGGTCCTTCCAGTGGGATCGAC 872

Db 148 TTCCACAGCTCGCGTTGAGGACAACTCTTCGGGTCCTTCCAGTGGGATCGGC 203

Search completed: February 19, 2003, 21:23:43

Job time : 583.739 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

nucleic - nucleic search, using sw model

on: February 19, 2003, 20:28:04 ; Search time 120.641 Seconds  
(without alignments)  
2399.710 Million cell updates/sec

file: US-09-808-388-7  
effect score: 944  
quence: 1 tgcggcctcgggtgagcc.....cgtcgtgactgggaaaccc 944

bring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

arched: 441362 seqs, 153338381 residues

al number of hits satisfying chosen parameters: 882724

nimum DB seq length: 0  
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

tabase : Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A COMB.seq: \*  
2: /cgn2\_6/prodata/2/ina/5B COMB.seq: \*  
3: /cgn2\_6/prodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/prodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/prodata/2/ina/PCUS COMB.seq: \*  
6: /cgn2\_6/prodata/2/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

sult No.	Score	Query Match	Length	DB	ID	Description
1	144	15.3	8083	4	US-09-383-630-4	Sequence 4, Appli
2	144	15.3	8083	4	US-09-383-630-5	Sequence 5, Appli
3	104.4	11.1	34303	2	US-08-735-609-4	Sequence 4, Appli
4	104.4	11.1	34303	2	US-08-735-609-4	Sequence 4, Appli
5	104.4	11.1	34303	3	US-09-315-372-4	Sequence 4, Appli
6	104.4	11.1	34303	3	US-09-244-752-4	Sequence 4, Appli
7	104.4	11.1	34303	3	US-09-245-497-4	Sequence 4, Appli
8	104.4	11.1	34303	4	US-09-562-919-4	Sequence 4, Appli
9	104.4	11.1	34382	2	US-08-374-483-6	Sequence 6, Appli
10	104.4	11.1	35408	4	US-08-973-334-3	Sequence 3, Appli
11	104.4	11.1	35408	4	US-09-563-869A-3	Sequence 3, Appli
12	104.4	11.1	35408	4	US-08-549-489-3	Sequence 3, Appli
13	104.4	11.1	35935	2	US-08-735-609-1	Sequence 1, Appli
14	104.4	11.1	35935	2	US-08-735-609-1	Sequence 1, Appli
15	104.4	11.1	35935	3	US-08-379-452-43	Sequence 43, Appli
16	104.4	11.1	35935	3	US-09-315-372-1	Sequence 1, Appli
17	104.4	11.1	35935	3	US-09-244-752-1	Sequence 1, Appli
18	104.4	11.1	35935	3	US-09-245-497-1	Sequence 1, Appli
19	104.4	11.1	35935	3	US-09-409-670-43	Sequence 43, Appli
20	104.4	11.1	35935	4	US-09-562-919-1	Sequence 1, Appli
21	98	10.4	343	5	PCT-US93-08067-1	Sequence 1, Appli
22	86.8	9.2	36519	3	US-08-923-137-2	Sequence 2, Appli
23	78.8	8.3	266	5	PCT-US93-08067-2	Sequence 2, Appli
24	69.6	7.4	3524	3	US-08-923-137-1	Sequence 1, Appli
25	64.8	6.9	35081	2	US-08-752-760A-1	Sequence 1, Appli
26	62	6.6	7218	1	US-08-232-463-14	Sequence 14, Appli
27	58	6.1	208	3	US-08-766-354A-1	Sequence 1, Appli

Sequence 7, Appli  
Sequence 11, Appli  
Sequence 3, Appli  
Sequence 14, Appli  
Sequence 7, Appli  
Sequence 13, Appli  
Sequence 8, Appli  
Sequence 11, Appli  
Sequence 12, Appli  
Sequence 15, Appli  
Sequence 14, Appli  
Sequence 16, Appli  
Sequence 14, Appli  
Sequence 23, Appli  
Sequence 27, Appli  
Sequence 28, Appli  
Sequence 55, Appli  
Sequence 56, Appli

## ALIGNMENTS

RESULT 1  
US-09-383-630-4  
; Sequence 4, Application US/09383630A  
; Patent No. 6265632  
; GENERAL INFORMATION:  
; APPLICANT: Avner Yavon et al.  
; TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH  
; FACTOR RECEPTOR ASSOCIATED  
; CHONDRODYSPLASIA  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Mark M. Friedman c/o Anthony Castorina  
; SFEET: 2001 Jefferson Davis Highway, Suite 207  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: United States of America  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
; COMPUTER: Twinhead\* Slimnote-890TX  
; OPERATING SYSTEM: MS DOS version 6.2,  
; SOFTWARE: Word for Windows version 2.0 converted  
; to an ASCII file  
; CURRENT APPLICATION NUMBER: US/09/383,630A  
; FILING DATE: 26-Aug-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: <Unknown>  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Friedman, Mark M.  
; REGISTRATION NUMBER: 33,893  
; REFERENCE/DOCKET NUMBER: 1402/2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 972-3-5625553  
; TELEFAX: 972-3-5625554  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8083  
; TYPE: nucleic acid  
; STRANDEDNESS: Double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-383-630-4

Query Match 15.3%; Score 144; DB 4; Length 8083;  
Best Local Similarity 96.7%; Pred. No. 3.4e-28;

Matches 147; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Y 732 TAGTGGATCCCCGGGCTGCAGATCTGTAGGCGCAGTAGTCCAGGGTTTCCTTGATGAT 791  
b 4867 TAGTGGATCCCCGGGCTGCAGATCCCCGGGCGCAGTAGTCCAGGGTTTCCTTGATGAT 4926

Y 792 GTCATACTTATCTCTGTCCTTTTTCACAGCTCGCGGTTGAGGACAAACTCTTCGCG 851  
b 4927 GTCATACTTATCTCTGTCCTTTTTCACAGCTCGCGGTTGAGGACAAACTCTTCGCG 4986

Y 852 GTCTTCCAGTGGGGATCGACGGTATCGATAA 883  
b 4987 GTCTTCCAGTGGGGATCGACGGTATCGATCA 5018

RESULT 2  
S-09-383-630-5  
Sequence 5, Application US/09383630A  
Patent No. 6265632  
GENERAL INFORMATION:  
APPLICANT: Avner Yayon et al.  
TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH  
FACTOR RECEPTOR ASSOCIATED  
CHONDRODYSPLASIA  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mark M. Friedman c/o Anthony Castorina  
STREET: 2001 Jefferson Davis Highway, Suite 207  
CITY: Arlington  
STATE: Virginia  
COUNTRY: United States of America  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
COMPUTER: Twinhead\* Slimnote-890TX  
OPERATING SYSTEM: MS DOS version 6.2,  
Windows version 3.11  
SOFTWARE: Word for Windows version 2.0 converted  
to an ASCII file  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/383,630A  
FILING DATE: 26-Aug-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: <Unknown>  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Friedman, Mark M.  
REGISTRATION NUMBER: 33,883  
REFERENCE/DOCKET NUMBER: 1402/2  
TELEPHONE: 972-3-5625553  
TELEFAX: 972-3-5625554  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8083  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
IS-09-383-630-5  
Query Match 15.3%; Score 144; DB 4; Length 8083;  
Best Local Similarity 96.7%; Pred. No. 3.4e-28;  
Matches 147; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Y 732 TAGTGGATCCCCGGGCTGCAGATCTGTAGGCGCAGTAGTCCAGGGTTTCCTTGATGAT 791  
b 4867 TAGTGGATCCCCGGGCTGCAGATCCCCGGGCGCAGTAGTCCAGGGTTTCCTTGATGAT 4926

Y 792 GTCATACTTATCTCTGTCCTTTTTCACAGCTCGCGGTTGAGGACAAACTCTTCGCG 851  
b 4927 GTCATACTTATCTCTGTCCTTTTTCACAGCTCGCGGTTGAGGACAAACTCTTCGCG 4986

Db 4927 GTCATACTTATCTCTGTCCTTTTTCACAGCTCGCGGTTGAGGACAAACTCTTCGCG 4986

QY 852 GTCTTCCAGTGGGGATCGACGGTATCGATAA 883  
Db 4987 GTCTTCCAGTGGGGATCGACGGTATCGATCA 5018

RESULT 3  
US-08-735-609-4  
Sequence 4, Application US/08735609  
Patent No. 5955360  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.  
APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.  
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/735,609  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: UM-02484  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
US-08-735-609-4  
Query Match 11.1%; Score 104.4; DB 2; Length 34303;  
Best Local Similarity 99.1%; Pred. No. 9.6e-18;  
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 757 TGTAGGGCGCAGTAGTCCAGGGTTCCCTTGATGATGATCATATCTGTCCTCTTTT 816  
Db 8546 TGTAGGGCGCAGTAGTCCAGGGTTCCCTTGATGATGATCATATCTGTCCTCTTTT 8605

QY 817 TTCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTTCTTCCAGT 862  
Db 8606 TTCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTTCTTCCAGT 8651

RESULT 4  
US-08-735-609-4  
Sequence 4, Application US/08735609  
Patent No. 594132  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.

Kumar-Singh, Rajendra  
Hartigan-O'Connor, Dennis J.  
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/735,609  
FILING DATE: 23-Oct-1996  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: UM-02484  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
-08-735-609-4  
Query Match 11.1%; Score 104.4; DB 2; Length 34303;  
Best Local Similarity 99.1%; Pred. No. 9.6e-18;  
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
757 TGTAGGGCGCAGTAGTCAGGGTTCCCTTGATGATGTCATCTATCCGTCTTCCAGT 816  
8546 TCTAGGGCGCAGTAGTCAGGGTTCCCTTGATGATGTCATCTATCCGTCTTCCAGT 8605  
817 TTCCACAGCTCGCGGTGAGGACAACTCTTCGCGTCTTCCAGT 862  
806 TTCCACAGCTCGCGGTGAGGACAACTCTTCGCGTCTTCCAGT 8651  
SULT 5  
-09-315-372-4  
Sequence 4, Application US/09315372  
Patent No. 6057158  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.  
APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.  
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/244,752  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/735,609  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/315,372  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/735,609  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
-09-315-372-4  
Query Match 11.1%; Score 104.4; DB 3; Length 34303;  
Best Local Similarity 99.1%; Pred. No. 9.6e-18;  
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
757 TGTAGGGCGCAGTAGTCAGGGTTCCCTTGATGATGTCATCTATCCGTCTTCCAGT 816  
8546 TCTAGGGCGCAGTAGTCAGGGTTCCCTTGATGATGTCATCTATCCGTCTTCCAGT 8605  
817 TTCCACAGCTCGCGGTGAGGACAACTCTTCGCGTCTTCCAGT 862  
806 TTCCACAGCTCGCGGTGAGGACAACTCTTCGCGTCTTCCAGT 8651  
RESULT 6  
US-09-244-752-4  
Sequence 4, Application US/09244752  
Patent No. 6063622  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Amalfitano, Andrea  
APPLICANT: Hauser, Michael A.  
APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.  
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/244,752  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/735,609  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: UM-02484

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 34303 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

IS-09-244-752-4

Query Match 11.1%; Score 104.4; DB 3; Length 34303;  
Best Local Similarity 99.1%; Pred. No. 9.6e-18;  
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 757 TGTAGGGCGCAGTAGTCACAGGTTTCCTTGATGATGTCATCTATCTGTCCTCTTTT 816

b 8546 TCTAGGGCGCAGTAGTCACAGGTTTCCTTGATGATGTCATCTATCTGTCCTCTTTT 8605

Y 817 TTCCACAGCTCGCGTTGAGGACAACTCTTCGGGTCCTTCCAGT 862

b 8606 TTCCACAGCTCGCGTTGAGGACAACTCTTCGGGTCCTTCCAGT 8651

RESULT 7

IS-09-245-497-4

Sequence 4, Application US/09245497

Patent No. 6083750

GENERAL INFORMATION:

APPLICANT: Chamberlain, Jeffrey S.

APPLICANT: Amalfitano, Andrea

APPLICANT: Hauser, Michael A.

APPLICANT: Kumar-Singh, Rajendra

APPLICANT: Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States Of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/245,497

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/735,609

APPLICATION NUMBER: <B> FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: UM-02484

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 34303 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-09-245-497-4

Query Match 11.1%; Score 104.4; DB 3; Length 34303;

Best Local Similarity 99.1%; Pred. No. 9.6e-18;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 757 TGTAGGGCGCAGTAGTCACAGGTTTCCTTGATGATGTCATCTATCTGTCCTCTTTT 816

Db 8546 TCTAGGGCGCAGTAGTCACAGGTTTCCTTGATGATGTCATCTATCTGTCCTCTTTT 8605

QY 817 TTCCACAGCTCGCGTTGAGGACAACTCTTCGGGTCCTTCCAGT 862

Db 8606 TTCCACAGCTCGCGTTGAGGACAACTCTTCGGGTCCTTCCAGT 8651

RESULT 8

US-09-562-919-4

Sequence 4, Application US/09562919

Patent No. 6451596

GENERAL INFORMATION:

APPLICANT: Chamberlain, Jeffrey S.

APPLICANT: Amalfitano, Andrea

APPLICANT: Hauser, Michael A.

APPLICANT: Kumar-Singh, Rajendra

APPLICANT: Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States Of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/562,919

FILING DATE: 02-May-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/735,609

FILING DATE: 23-Oct-1996

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: UM-02484

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 34303 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

US-09-562-919-4

Query Match 11.1%; Score 104.4; DB 4; Length 34303;

Best Local Similarity 99.1%; Pred. No. 9.6e-18;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 757 TGTAGGGCGCAGTAGTCACAGGTTTCCTTGATGATGTCATCTATCTGTCCTCTTTT 816

Db 8546 TCTAGGGCGCAGTAGTCACAGGTTTCCTTGATGATGTCATCTATCTGTCCTCTTTT 8605

QY 817 TTCCACAGCTCGCGTTGAGGACAACTCTTCGGGTCCTTCCAGT 862

8606 TTCCACAGCTCGCGGTGAGGACAACTCTTCGGGTCCTTCCAGT 8651

SULT 9

-08-374-483-6  
Sequence 6, Application US/08374483  
Patent No. 5880102

GENERAL INFORMATION:

APPLICANT: GEORGE, SAMUEL E.

APPLICANT: BLAZING, MICHAEL A.

TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/374,483

FILING DATE: 17-JAN-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: WILSON, MARY J.

REGISTRATION NUMBER: 32,955

REFERENCE/DOCKET NUMBER: 1579-83

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 34382 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

-08-374-483-6

Query Match 11.1%; Score 104.4; DB 2; Length 34382;  
Best Local Similarity 99.1%; Pred. No. 9.6e-18;  
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATCCCTGCTCTTTT 816

7966 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATCCCTGCTCTTTT 8025

817 TTCCACAGCTCGCGGTGAGGACAACTCTTCGGGTCCTTCCAGT 862

8026 TTCCACAGCTCGCGGTGAGGACAACTCTTCGGGTCCTTCCAGT 8071

SULT 10

-08-973-334-3

Sequence 3, Application US/08973334

Patent No. 6261551

GENERAL INFORMATION:

APPLICANT: Wilson, James M.

APPLICANT: Fisher, Krishna J.

APPLICANT: Gao, Guang-Ping

TITLE OF INVENTION: Recombinant Adenovirus and Adeno-

TITLE OF INVENTION: Associated Virus, Cell Lines, and

TITLE OF INVENTION: Methods of Production and Use

TITLE OF INVENTION: Thereof

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson  
STREET: Box 457, 321 No. 6261551ristown Road

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release 1.0 Version 1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/973,334

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/462,014

FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/549,489

FILING DATE: 27-OCT-1995

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: GNPVN012CIPUSA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 540-9206

TELEFAX: (215) 540-5818

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 35408 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: other nucleic acid

US-08-973-334-3

Query Match 11.1%; Score 104.4; DB 4; Length 35408;

Best Local Similarity 99.1%; Pred. No. 9.6e-18;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATCCCTGCTCTTTT 816

Db 8371 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATCCCTGCTCTTTT 8430

Qy 817 TTCCACAGCTCGCGGTGAGGACAACTCTTCGGGTCCTTCCAGT 862

Db 8431 TTCCACAGCTCGCGGTGAGGACAACTCTTCGGGTCCTTCCAGT 8476

RESULT 11

US-09-563-869A-3

Sequence 3, Application US/09563869A

Patent No. 6270996

GENERAL INFORMATION:

APPLICANT: Wilson, James M.

APPLICANT: Fisher, Krishna J.

APPLICANT: Gao, Guang-Ping

TITLE OF INVENTION: Recombinant Adenovirus and Adeno-

TITLE OF INVENTION: Associated Virus, Cell Lines, and

TITLE OF INVENTION: Methods of Production and Use

TITLE OF INVENTION: Thereof

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Box 457, 321 No. 6270996ristown Road

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release 1.0 Version 1.30  
CURRENT APPLICATION DATA: US/09/563,869A  
FILING DATE: 03-May-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/973,334  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/549,489  
FILING DATE: 27-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNVN012CIPUSA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 540-9206  
TELEFAX: (215) 540-5818  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35408 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
IS-09-563-869A-3  
Query Match 11.1%; Score 104.4; DB 4; Length 35408;  
Best Local Similarity 99.1%; Pred. No. 9.6e-18;  
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Y 757 TGTAGGGCGAGTAGTCACGGGTTTCCTTGATGATGCATATCCTGTCCTCTTTT 816  
b 8371 TCTAGGGCGAGTAGTCACGGGTTTCCTTGATGATGCATATCCTGTCCTCTTTT 8430  
Y 817 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 862  
b 8431 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 8476  
RESULT 12  
US-08-549-489-3  
Sequence 3, Application US/08549489  
Patent No. 6281010  
GENERAL INFORMATION:  
APPLICANT: Wilson, James M.  
APPLICANT: Gao, Guang-Ping  
TITLE OF INVENTION: No. 6281010el Adenovirus Gene Therapy Vehicle  
TITLE OF INVENTION: and Cell Line  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Box 457, 321 No. 6281010ristown Road  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/549,489  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/462,014  
FILING DATE: 08-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: GNVN013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 540-9206  
TELEFAX: (215) 540-5818  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35408 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: other nucleic acid  
US-08-549-489-3  
Query Match 11.1%; Score 104.4; DB 4; Length 35408;  
Best Local Similarity 99.1%; Pred. No. 9.6e-18;  
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 757 TGTAGGGCGAGTAGTCACGGGTTTCCTTGATGATGCATATCCTGTCCTCTTTT 816  
Db 8371 TCTAGGGCGAGTAGTCACGGGTTTCCTTGATGATGCATATCCTGTCCTCTTTT 8430  
QY 817 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 862  
Db 8431 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 8476  
RESULT 13  
US-08-735-609-1  
Sequence 1, Application US/08735609  
Patent No. 5955360  
GENERAL INFORMATION:  
APPLICANT: Chamberlain, Jeffrey S.  
APPLICANT: Analfitano, Andrea  
APPLICANT: Hauser, Michael A.  
APPLICANT: Kumar-Singh, Rajendra  
APPLICANT: Hartigan-O'Connor, Dennis J.  
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/735,609  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: UM-02484  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35935 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
US-08-735-609-1  
Query Match 11.1%; Score 104.4; DB 2; Length 35935;  
Best Local Similarity 99.1%; Pred. No. 9.7e-18;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

757 TGTAGGGCGAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATTCCTGTCCTCTTTT 816  
7043 TCTAGGGCGAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATTCCTGTCCTTTT 7102  
817 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 862  
7103 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 7148

SULT 14

-08-735-609-1

Sequence 1, Application US/08735609

Patent No. 5994132

GENERAL INFORMATION:

APPLICANT: Chamberlain, Jeffrey S.

Amalfitano, Andrea

Hauser, Michael A.

Kumar-Singh, Rajendra

Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States Of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/735,609

FILING DATE: 23-Oct-1996

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: UM-02484

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 35935 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

-08-735-609-1

Query Match 11.1%; Score 104.4; DB 2; Length 35935;

Best Local Similarity 99.1%; Pred. No. 9.7e-18;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

757 TGTAGGGCGAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATTCCTGTCCTTTT 816  
7043 TCTAGGGCGAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATTCCTGTCCTTTT 7102  
817 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 862  
7103 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 7148

SULT 15

-08-379-452-43

Sequence 43, Application US/08379452

Patent No. 6040174  
GENERAL INFORMATION:  
APPLICANT: IMLER, Jean-Luc  
APPLICANT: MEHTALI, Majid  
APPLICANT: PAVIRANI, Andrea  
TITLE OF INVENTION: DEFECTIVE ADENOVIRUSES AND CORRESPONDING  
TITLE OF INVENTION: COMPLEMENTATION LINES  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: BURNS, DOANE, SWICKER & MATHIS, L.L.P.  
STREET: 1737 King Street, Suite 500  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22314-2756  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/379,452  
FILING DATE: 26-JAN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR94/00624  
FILING DATE: 27-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 93 06482  
FILING DATE: 28-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Dadio, Susan M.  
REGISTRATION NUMBER: 40,373  
REFERENCE/DOCKET NUMBER: 029395-002  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35935 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

US-08-379-452-43

Query Match 11.1%; Score 104.4; DB 3; Length 35935;

Best Local Similarity 99.1%; Pred. No. 9.7e-18;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 757 TGTAGGGCGAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATTCCTGTCCTTTT 816  
Db 7043 TCTAGGGCGAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTATTCCTGTCCTTTT 7102  
Qy 817 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 862  
Db 7103 TTCCACAGCTCGCGGTTGAGGACAACTCTTCGGGTCCTTCCAGT 7148

Search completed: February 20, 2003, 01:19:13

Job time : 279.641 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

A nucleic - nucleic search, using sw model

in on: February 19, 2003, 23:17:14 ; Search time 716.061 Seconds  
(without alignments)  
671.453 Million cell updates/sec

file: US-09-808-388-7

arfect score: 944

sequence: 1 tgcggcctcggtgagcc.....cgctgactgggaaaccc 944

oring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

arched: 424239 seqs, 254661826 residues

tal number of hits satisfying chosen parameters: 848478

imum DB seq length: 0

ximum DB seq length: 200000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	944	100.0	944	10	US-09-808-388-7
2	104.4	11.1	1240	9	US-09-847-101B-28
3	104.4	11.1	7231	9	US-09-847-101B-42
4	104.4	11.1	7960	9	US-09-847-101B-30
5	104.4	11.1	7989	9	US-09-847-101B-33
6	104.4	11.1	8383	9	US-09-847-101B-29
7	104.4	11.1	32480	9	US-09-847-101B-23
8	104.4	11.1	35871	10	US-09-956-335-2
9	104.4	11.1	35935	10	US-09-725-720-43
10	104.4	11.1	35935	10	US-09-782-378A-4
11	104.4	11.1	35935	10	US-09-782-378A-5
12	104.4	11.1	35937	10	US-09-782-378A-3
13	104.4	11.1	35978	10	US-09-956-335-1
14	64.8	6.9	35100	10	US-09-782-378A-26
15	58.6	6.2	34214	10	US-09-782-378A-27
16	54.8	5.8	12733	9	US-10-032-393-47
17	54.8	5.8	12739	9	US-10-032-393-8
18	53.8	5.7	1065	10	US-09-804-682-33
19	48.6	5.1	987	10	US-09-804-682-20

20	48.2	5.1	34125	10	US-09-782-378A-25
21	45.6	4.8	1064	10	US-09-804-682-29
c 22	44.6	4.7	457	10	US-09-867-701-3922
23	44.2	4.7	504	10	US-09-878-574-4330
c 24	44	4.7	106	10	US-09-960-352-5165
c 25	43.4	4.6	366	10	US-09-878-574-4418
c 26	43	4.6	616	10	US-09-878-574-4567
27	42.6	4.5	152331	9	US-10-095-407-16
28	41	4.3	41	9	US-10-158-735-7
29	40.8	4.3	266	10	US-09-960-352-3882
30	40	4.2	3008	10	US-09-764-855-298
31	39.6	4.2	4279	10	US-09-956-998A-1
c 32	39.6	4.2	4950	10	US-09-916-940-58
c 33	39.6	4.2	4950	10	US-09-918-601-58
c 34	39.6	4.2	6822	10	US-09-836-077-42
35	39	4.1	327	9	US-09-847-101B-22
36	39	4.1	3788	10	US-09-392-462-1
37	39	4.1	3788	10	US-09-540-991-1
38	39	4.1	5155	9	US-10-128-853-3
39	39	4.1	5374	9	US-10-007-357-7
40	39	4.1	5462	9	US-10-007-357-6
41	39	4.1	7469	9	US-09-847-101B-8
42	39	4.1	10610	9	US-09-847-101B-16
c 43	39	4.1	14455	9	US-09-847-101B-15
44	38.8	4.1	2288	9	US-09-989-920-103
c 45	38.8	4.1	3662	9	US-10-001-189-41

ALIGNMENTS

RESULT 1

US-09-808-388-7  
; Sequence 7, Application US/09808388  
; Patent No. US20020081719A1

; GENERAL INFORMATION:  
; APPLICANT: Massaad, Charbel  
; APPLICANT: Berenbaum, Francis  
; APPLICANT: Olivier, Jean-Luc  
; APPLICANT: Salvat, Colette  
; APPLICANT: Berezziat, Gilbert  
; TITLE OF INVENTION: Inflammation Inducible Hybrid Promoters, Vectors Comprising them  
; TITLE OF INVENTION: their uses  
; FILE REFERENCE: ST00010  
; CURRENT APPLICATION NUMBER: US/09/808,388  
; PRIOR FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: FR/00/03262  
; PRIOR FILING DATE: 2000-03-14  
; PRIOR APPLICATION NUMBER: US 60/196,959  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 7

; LENGTH: 944  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Sequence conferring specificity of expression

US-09-808-388-7

Query Match 100.0%; Score 944; DB 10; Length 944;

Best Local Similarity 100.0%; Pred. No. 1.1e-240;

Mismatches 0; Indels 0; Gaps 0;

Matches 944; Conservative 0;

QY 1 TCCCGGCTCGGGTACCTGATCGCCTCGGGGTCGCCAGTCCCTGCTGCTGCTGA 60

Db 1 TCCCGGCTCGGGTACCTGATCGCCTCGGGGTCGCCAGTCCCTGCTGCTGCTGA 60

QY 61 CGCTCTCATCGCGCGTCTACGGTGTAGGCGCCAGGATGCCGTAAGTCGCCGCCG 120

Db 61 CGCTCTCATCGCGCGTCTACGGTGTAGGCGCCAGGATGCCGTAAGTCGCCGCCG 120

QY 121 CCCCTGCTACTTCCCTGACTGTGACCCCTTTCTCTACTCTCCCTCCCAAGTACTAG 180



```
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: plasmid pDV67
3-09-847-101B-30

Query Match      11.1%; Score 104.4; DB 9; Length 7960;
Best Local Similarity 99.1%; Pred. No. 4.1e-18;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTTATCCTGTCCTCTTTT 816
1929 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTTATCCTGTCCTCTTTT 1988

817 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGT 862
1989 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGT 2034

SULT 5
3-09-847-101B-33
Sequence 33, Application US/09847101B
Publication No. US20020193327A1
GENERAL INFORMATION:
APPLICANT: VON SEGGERN, DANIEL
APPLICANT: NEMEROW, GLEN R.
APPLICANT: FRIEDLANDER, MARTIN
TITLE OF INVENTION: VECTORS FOR OCULAR TRANSDUCTION AND USE THEREFOR FOR GENETIC THER
FILE REFERENCE: 22908-1226B
CURRENT APPLICATION NUMBER: US/09/847,101B
CURRENT FILING DATE: 2001-05-01
PRIOR APPLICATION NUMBER: 09/562,934
PRIOR FILING DATE: 2000-05-01
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 33
LENGTH: 7989
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4242
OTHER INFORMATION: N is any
NAME/KEY: misc_feature
LOCATION: 4245
OTHER INFORMATION: N is any
OTHER INFORMATION: Description of Artificial Sequence: plasmid pDV69
3-09-847-101B-33

Query Match      11.1%; Score 104.4; DB 9; Length 7989;
Best Local Similarity 99.1%; Pred. No. 4.1e-18;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTTATCCTGTCCTCTTTT 816
1929 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTTATCCTGTCCTCTTTT 1988

817 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGT 862
1989 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGT 2034

SULT 6
3-09-847-101B-29
Sequence 29, Application US/09847101B
Publication No. US20020193327A1
GENERAL INFORMATION:
APPLICANT: VON SEGGERN, DANIEL
APPLICANT: NEMEROW, GLEN R.
APPLICANT: FRIEDLANDER, MARTIN
TITLE OF INVENTION: VECTORS FOR OCULAR TRANSDUCTION AND USE THEREFOR FOR GENETIC THER
FILE REFERENCE: 22908-1226B
CURRENT APPLICATION NUMBER: US/09/847,101B
```

```
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/562,934
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 29
; LENGTH: 8383
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pDV60
US-09-847-101B-29

Query Match      11.1%; Score 104.4; DB 9; Length 8383;
Best Local Similarity 99.1%; Pred. No. 4.2e-18;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTTATCCTGTCCTCTTTT 816
1907 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTTATCCTGTCCTCTTTT 1966

817 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGT 862
1967 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGT 2012

RESULT 7
US-09-847-101B-23
; Sequence 23, Application US/09847101B
; Publication No. US20020193327A1
; GENERAL INFORMATION:
; APPLICANT: VON SEGGERN, DANIEL
; APPLICANT: NEMEROW, GLEN R.
; APPLICANT: FRIEDLANDER, MARTIN
; TITLE OF INVENTION: VECTORS FOR OCULAR TRANSDUCTION AND USE THEREFOR FOR GENETIC THER
; FILE REFERENCE: 22908-1226B
; CURRENT APPLICATION NUMBER: US/09/847,101B
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/562,934
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 23
; LENGTH: 32480
; TYPE: DNA
; ORGANISM: adenovirus
US-09-847-101B-23

Query Match      11.1%; Score 104.4; DB 9; Length 32480;
Best Local Similarity 99.1%; Pred. No. 6.7e-18;
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

757 TGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTTATCCTGTCCTCTTTT 816
8187 TCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACCTTATCCTGTCCTCTTTT 8246

817 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGT 862
8247 TTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGT 8292

RESULT 8
US-09-956-335-2
; Sequence 2, Application US/09956335
; Patent No. US20020028785A1
; GENERAL INFORMATION:
; APPLICANT: WOLD, William
; APPLICANT: TOFF, Karoly
; APPLICANT: KUPPASWAMI, Mohan
; APPLICANT: DORONIN, Konsantin
; TITLE OF INVENTION: RECOMBINANT ADENOVIRUS VECTORS THAT ARE
; TITLE OF INVENTION: REPLICATION-COMPETENT IN TERT-EXPRESSING CELLS
; FILE REFERENCE: 16153-8394
```

CURRENT APPLICATION NUMBER: US/09/956,335

CURRENT FILING DATE: 2001-09-19

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 35871

TYPE: DNA

ORGANISM: Adenovirus

US-09-956-335-2

Query Match 11.1%; Score 104.4; DB 10; Length 35871;  
Best Local Similarity 99.1%; Pred. No. 6.9e-18;  
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 757 TGTAGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTTATCCTGTCCTTTTTT 816

b 6938 TCTAGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTTATCCTGTCCTTTTTT 6997

Y 817 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTCCAGT 862

b 6998 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTCCAGT 7043

#### RESULT 9

US-09-725-720-43

Sequence 43, Application US/09725720

Patent No. US20010049136A1

GENERAL INFORMATION:

APPLICANT: IMLER, Jean-Luc

APPLICANT: MEHTALI, Majid

APPLICANT: PAVIRANI, Andrea

TITLE OF INVENTION: DEFECTIVE ADENOVIRUSES AND CORRESPONDING

TITLE OF INVENTION: COMPLEMENTATION LINES

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: 1737 King Street, Suite 500

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22314-2756

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/725,720

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/379,452

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 93 06482

FILING DATE: 28-MAY-1993

ATTORNEY/AGENT INFORMATION:

NAME: Dadio, Susan M.

REGISTRATION NUMBER: 40,373

REFERENCE/DOCKET NUMBER: 029395-002

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 35935 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-725-720-43

Query Match 11.1%; Score 104.4; DB 10; Length 35935;  
Best Local Similarity 99.1%; Pred. No. 6.9e-18;  
Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 757 TGTAGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTTATCCTGTCCTTTTTT 816

Db 7043 TCTAGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTTATCCTGTCCTTTTTT 7102

QY 817 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTCCAGT 862

Db 7103 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTCCAGT 7148

#### RESULT 10

US-09-782-378A-4

Sequence 4, Application US/09782378A

Patent No. US20020102731A1

GENERAL INFORMATION:

APPLICANT: Hearing, Patrick

APPLICANT: Bahou, Wadie

APPLICANT: Sandalon, Ziv

APPLICANT: Gnatenko, Dmitri

TITLE OF INVENTION: Adenoviral Vectors

FILE REFERENCE: STONYB-04970

CURRENT APPLICATION NUMBER: US/09/782,378A

CURRENT FILING DATE: 2001-02-12

PRIOR APPLICATION NUMBER: 60/237,747

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn version 3.0

SEQ ID NO 4

LENGTH: 35935

TYPE: DNA

ORGANISM: Human adenovirus type 5

US-09-782-378A-4

Query Match 11.1%; Score 104.4; DB 10; Length 35935;

Best Local Similarity 99.1%; Pred. No. 6.9e-18;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 757 TGTAGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTTATCCTGTCCTTTTTT 816

Db 7043 TCTAGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTTATCCTGTCCTTTTTT 7102

QY 817 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTCCAGT 862

Db 7103 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTCCAGT 7148

#### RESULT 11

US-09-782-378A-5

Sequence 5, Application US/09782378A

Patent No. US20020102731A1

GENERAL INFORMATION:

APPLICANT: Hearing, Patrick

APPLICANT: Bahou, Wadie

APPLICANT: Sandalon, Ziv

APPLICANT: Gnatenko, Dmitri

TITLE OF INVENTION: Adenoviral Vectors

FILE REFERENCE: STONYB-04970

CURRENT APPLICATION NUMBER: US/09/782,378A

CURRENT FILING DATE: 2001-02-12

PRIOR APPLICATION NUMBER: 60/237,747

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn version 3.0

SEQ ID NO 5

LENGTH: 35935

TYPE: DNA

ORGANISM: Human adenovirus type 5

US-09-782-378A-5

Query Match 11.1%; Score 104.4; DB 10; Length 35935;

Best Local Similarity 99.1%; Pred. No. 6.9e-18;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 757 TGTAGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTTATCCTGTCCTTTTTT 816

7043 TCTAGGGCGCAGTAGTCAGGGTTTCCTTGAAGTGCATACCTATCCCTGTCCTTTT 7102

817 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTTCCAGT 862

7103 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTTCCAGT 7148

## SULT 12

-09-782-378A-3

Sequence 3, Application US/09782378A

Patent No. US20020102731A1

GENERAL INFORMATION:

APPLICANT: Hearing, Patrick

APPLICANT: Bahou, Wadie

APPLICANT: Sandalon, Ziv

APPLICANT: Gnatenko, Dmitri

TITLE OF INVENTION: Adenoviral Vectors

FILE REFERENCE: STONYB-04970

CURRENT APPLICATION NUMBER: US/09/782,378A

CURRENT FILING DATE: 2001-02-12

PRIOR APPLICATION NUMBER: 60/237,747

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn version 3.0

SEQ ID NO 3

LENGTH: 35937

TYPE: DNA

ORGANISM: Human adeno-associated virus 2

-09-782-378A-3

Query Match

Best Local Similarity 11.1%; Score 104.4; DB 10; Length 35937;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

757 TGTAGGGCGCAGTAGTCAGGGTTTCCTTGAATGTCATACCTATCCCTGTCCTTTT 816

7033 TCTAGGGCGCAGTAGTCAGGGTTTCCTTGAATGTCATACCTATCCCTGTCCTTTT 7092

817 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTTCCAGT 862

7093 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTTCCAGT 7138

## SULT 13

-09-956-335-1

Sequence 1, Application US/09956335

Patent No. US2002028785A1

GENERAL INFORMATION:

APPLICANT: WOULD, William

APPLICANT: TOTTH, Karoly

APPLICANT: KUPPASWAMI, Mohan

APPLICANT: DORONIN, Konsantin

TITLE OF INVENTION: RECOMBINANT ADENOVIRUS VECTORS THAT ARE

FILE REFERENCE: 16153-8394

CURRENT APPLICATION NUMBER: US/09/956,335

CURRENT FILING DATE: 2001-09-19

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 35978

TYPE: DNA

ORGANISM: Adenovirus

-09-956-335-1

Query Match

Best Local Similarity 11.1%; Score 104.4; DB 10; Length 35978;

Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

757 TGTAGGGCGCAGTAGTCAGGGTTTCCTTGAATGTCATACCTATCCCTGTCCTTTT 816

7043 TCTAGGGCGCAGTAGTCAGGGTTTCCTTGAATGTCATACCTATCCCTGTCCTTTT 7102

Qy 817 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTTCCAGT 862

Db 7103 TTCCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTTCCAGT 7148

## RESULT 14

US-09-782-378A-26

Sequence 26, Application US/09782378A

Patent No. US20020102731A1

GENERAL INFORMATION:

APPLICANT: Hearing, Patrick

APPLICANT: Bahou, Wadie

APPLICANT: Sandalon, Ziv

APPLICANT: Gnatenko, Dmitri

TITLE OF INVENTION: Adenoviral Vectors

FILE REFERENCE: STONYB-04970

CURRENT APPLICATION NUMBER: US/09/782,378A

CURRENT FILING DATE: 2001-02-12

PRIOR APPLICATION NUMBER: 60/237,747

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn version 3.0

SEQ ID NO 26

LENGTH: 35100

TYPE: DNA

ORGANISM: Human adenovirus type 17

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (25184)..(25184)

OTHER INFORMATION: The n at this position can be a, c, t, or g.

US-09-782-378A-26

Query Match

Best Local Similarity 6.9%; Score 64.8; DB 10; Length 35100;

Matches 87; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 739 TCCGCCGGGTGCAGATCTGTAGGGCGCAGTAGTCAGGGTTTCCTTGAATGTCATAC 798

Db 6819 TCGGCGGTGACGAGACGTCCTCATGGCGCAGTAGTCAGGGTTTCGCGGATGATGCATAA 6878

Qy 799 TTATCCTGTCCTTTTTCACAGCTCGCGTTGAGGACAAACTCTTCGGGTCTTTTC 858

Db 6879 CCGCCTCTCCTTTCTTCTCCACAGCTCGCGTTGAGGCGGTATTCTCTGTCATCCTTC 6938

Qy 859 CAGT 862

Db 6939 CAGT 6942

## RESULT 15

US-09-782-378A-27

Sequence 27, Application US/09782378A

Patent No. US20020102731A1

GENERAL INFORMATION:

APPLICANT: Hearing, Patrick

APPLICANT: Bahou, Wadie

APPLICANT: Sandalon, Ziv

APPLICANT: Gnatenko, Dmitri

TITLE OF INVENTION: Adenoviral Vectors

FILE REFERENCE: STONYB-04970

CURRENT APPLICATION NUMBER: US/09/782,378A

CURRENT FILING DATE: 2001-02-12

PRIOR APPLICATION NUMBER: 60/237,747

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn version 3.0

SEQ ID NO 27

LENGTH: 34214

TYPE: DNA

ORGANISM: Human adenovirus type 40

US-09-782-378A-27

Query Match 6.2%; Score 58.6; DB 10; Length 34214;  
Best Local Similarity 65.9%; Pred. NO. 9.1e-06;  
Matches 85; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
Y 760 AGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATCTATCTATCTGTCCTGTCCTTTTTC 819  
b 6596 AGGGCACAGTAATCAAGGGTTTCACGCCACCGAGTCGTATGAGGACATTGCTTTTTC 6655  
Y 820 CACAGCTCGCGGTGAGGACAAACTCTTCGGGTCTTTCCAGTGGGGATCGACGGTATCG 879  
b 6656 CAGAGTTCGGCGGTTCAGGAGGTACTCTCCGATCCTCCAGTAATCTTCGGCAGGAAAG 6715  
Y 880 ATRAGCTTG 888  
b 6716 CCACGCTCG 6724

Search completed: February 20, 2003, 07:03:49  
Job time : 829.061 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

[ nucleic - nucleic search, using sw model

in on: February 19, 2003, 19:33:48 ; Search time 4593.24 Seconds  
(without alignments)  
3328.484 Million cell updates/sec

tle: US-09-808-388-7  
rfect score: 944  
quence: 1 tgcggcctcgcgtagcc.....cgtcgtagtgggaaaccc 944

oring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

arched: 16154066 seqs, 8097743376 residues

tal number of hits satisfying chosen parameters: 32308132

nimum DB seq length: 0  
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase :

BST:\*

- 1: em\_estba:\*
- 2: em\_esthm:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_prt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	173.2	18.3	631	10	BB665483
2	131.4	13.9	285	17	AZ827835
3	102.6	10.9	879	14	BQ714997
4	102.6	10.9	882	14	BQ712993
5	102.6	10.9	894	14	BQ712860
6	102.6	10.9	915	14	BQ955920

7	102.6	10.9	941	14	BQ934062	
8	102.6	10.9	962	14	BQ718103	
9	102.6	10.9	965	14	BQ715834	
10	102.6	10.9	976	14	BQ950196	
11	102.6	10.9	976	14	BQ956014	
12	102.6	10.9	1003	14	BQ714350	
13	102.6	10.9	1024	14	BQ934161	
14	102.6	10.9	1051	14	BQ714753	
15	102.6	10.8	477	14	B78398	
16	100.4	10.6	442	10	B839890	
17	100.4	10.6	445	10	B839999	
18	100.4	10.6	447	10	B840435	
19	100.4	10.6	454	10	B840198	
20	100.4	10.6	474	10	B839913	
21	99.4	10.5	539	17	BQ065083	
22	98.8	10.5	458	10	B840209	
23	97.2	10.3	1076	14	BQ917370	
c	24	87	565	10	AV592867	
	25	86.6	9.2	1116	14	BQ715301
	26	79.8	8.5	466	10	BQ839921
	27	78.6	8.3	953	17	CNS017Y4
	28	72.2	7.6	196	14	W13732
29	71.6	7.6	351	9	AA168904	
30	71	7.5	490	10	BQ523745	
c	31	70.4	7.5	988	17	AG135153
	32	70.2	7.4	908	17	AG137973
c	33	70	7.4	1104	14	BQ883653
	34	69.2	7.3	511	9	AL514069
c	35	68.8	7.3	785	17	AG136803
	36	68.8	7.3	802	17	CNS02343
c	37	68	7.2	372	13	B1885632
	38	68	7.2	1247	14	BM809415
c	39	67.6	7.2	745	17	AG046015
	40	67.6	7.2	809	17	AG136883
c	41	67.4	7.1	511	9	AJ437840
	42	67.2	7.1	873	17	AG132025
c	43	67.2	7.1	916	17	AG137890
	44	67	7.1	473	12	BQ519370
45	66.8	7.1	424	12	BQ655165	

ALIGNMENTS

RESULT 1  
BB665483  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BB665483 631 bp mRNA linear EST 26-OCT-2001  
BB665483 RIKEN full-length enriched, 0 day neonate eyeball Mus  
musculus cDNA clone EI30308G17 5', mRNA sequence.  
BB665483  
EST.  
BB665483.1 GI:16499116  
Mus musculus  
House mouse.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 631)  
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,  
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Kono, H., Kouda  
M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ono, M., Sasaki  
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki  
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,  
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,  
Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
 Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
 Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa  
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
 Hayashizaki, Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details  
 e mouse tissues. Location/Qualifiers

## FEATURES

source  
 1. 631  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="E130308G17"  
 /clone\_lib="RIKEN full-length enriched, 0 day neonate  
 eyeball"  
 /tissue\_type="eyeball"  
 /dev\_stage="0 day neonate"  
 /lab\_host="DH10B"  
 /notes="Site 1: SalI; Site 2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGCGCGCGCACTCGAGTCTTTTCTTTTCTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. Second strand cDNA was prepared with the  
 primer adapter of sequence [5'  
 GAGAGAGATTCTCGAGTTAATTAATTAATTCCTCCCTCCCGCCG 3']. cDNA  
 was cleaved with BamHI and XhoI. Vector: a modified  
 pBluescript KS(+) after bulk excision from Lambda PLC I."  
 119 a 202 c 177 g 133 t

## BASE COUNT

ORIGIN

Query Match 18.3%; Score 173.2; DB 10; Length 631;  
 Best Local Similarity 98.3%; Pred. No. 1.2e-29;  
 Matches 175; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 2y 3 CCGGCTCGCGTGAGCCCTGATCGCTCGGGGCTCCCGAGTCGCTGCTGCTGAGC 62  
 db 165 CAGGCTCGCGTGAGCCATGATCGCTCGGGGCTCCCGAGTCGCTGCTGCTGAGC 224  
 2y 63 CTGCTCATCGCGGGTCTCTAGGTGTGAGGGCAGGATGCCGTAGTCCCGCCGCGCC 122  
 db 225 CTGCTCATCGCGGGTCTCTAGGTGTGAGGGCAGGATGCCGTAGTCCCGCCGCGCC 284  
 2y 123 CTGCTCATCTCCCTGACTTGACCTTTTCTCTACTCCCTCCCGCAAGTACTAG 180  
 db 285 CTTGCTACTTCCCTGACTTGACCTTTTCTCTACTCCCTCCCGCAAGTACTAG 342

## RESULT 2

AZ827835 285 bp DNA linear GSS 20-FEB-2001  
 LOCUS 2M0104E01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 DEFINITION clone UUGC2M0104E01 R, DNA sequence.  
 ACCESSION AZ827835  
 VERSION AZ827835.1 GI:12997743  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 REFERENCE 1 (bases 1 to 285)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D. Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 UNPUBLISHED (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0104 row: E column: 01  
 Seq primer: CACACAGAAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 285.  
 Location/Qualifiers

## FEATURES

source  
 1. 285  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0104E01"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (GI:4732114|GB|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."  
 55 a 74 c 79 g 77 t

## BASE COUNT

ORIGIN

Query Match 13.9%; Score 131.4; DB 17; Length 285;  
 Best Local Similarity 87.3%; Pred. No. 4.3e-20;  
 Matches 144; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
 Qy 410 CCGCTGCCCGCTGCCGCCACCTTCAGATCGATCTGGGATTGCGAGCGATGCTTCCAGA 469  
 Db 121 CCTCAGTCCTCTTTGTGAGACTGTTCGTTGGGGATTGGCAGCGATGCTTCCAGA 180  
 Qy 470 TGGCTGAAACCTGCCCGCTATTATTAACTGGTCTCTGTGGAGACTGTGAATCGG 529



```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6307602"
/clone_lib="NIH MGC 129"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dr. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH MGC Library."
ASE COUNT 164 a 304 c 309 g 117 t
RIGIN

Query Match 10.9%; Score 102.6; DB 14; Length 894;
Best Local Similarity 96.3%; Pred. No. 2.1e-13;
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Y 3 CCGGCTCGCGGTGAGCCCTGATCCGCTCGGGGTCCTCCAGTCCGCTGCTGCTGACG 62
b 140 CAGGCTCGCGGTGAGCCATGATCCGCTCGGGGTCCTCCATGCTGCTGCTGACG 199

Y 63 CTGCTCATCCGCGGTCTTACGGTGTACGGGTCAGGCGCAGGATGCCGTAAGT 111
b 200 CTGCTCATCCGCGGTCTTACGGTGTACGGGTCAGGCGCAGGATGCCGTAAGT 248

RESULT 6
LOCUS BQ955920 915 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8801763 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6315379
5', mRNA sequence.
ACCESSION BQ955920
VERSION BQ955920.1 GI:22371398
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
CDNA Library Preparation: ResGen, Invitrogen Corp
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13745 row: j column: 20
High quality sequence stop: 652.
Location/Qualifiers
1..915
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6315379"
/clone_lib="NIH MGC 129"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dr. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH MGC Library."
ASE COUNT 166 a 280 c 325 g 143 t 1 others
RIGIN

Query Match 10.9%; Score 102.6; DB 14; Length 915;
Best Local Similarity 96.3%; Pred. No. 2.1e-13;
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Y 3 CCGGCTCGCGGTGAGCCCTGATCCGCTCGGGGTCCTCCAGTCCGCTGCTGCTGACG 62

```

```

Db 119 CAGGCTCGCGGTGAGCCATGATCCGCTCGGGGTCCTCCAGTCCGCTGCTGCTGACG 178
QY 63 CTGCTCATCCGCGGTCTTACGGTGTACGGGTCAGGCGCAGGATGCCGTAAGT 111
Db 179 CTGCTCATCCGCGGTCTTACGGTGTACGGGTCAGGCGCAGGATGCCGTAAGT 227

RESULT 7
LOCUS BQ934062 941 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8764844 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6314533
5', mRNA sequence.
ACCESSION BQ934062
VERSION BQ934062.1 GI:22349445
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
CDNA Library Preparation: ResGen, Invitrogen Corp
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13743 row: g column: 14
High quality sequence stop: 556.
Location/Qualifiers
1..941
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6314533"
/clone_lib="NIH MGC 129"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dr. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH MGC Library."
BASE COUNT 173 a 291 c 323 g 154 t
ORIGIN

Query Match 10.9%; Score 102.6; DB 14; Length 941;
Best Local Similarity 96.3%; Pred. No. 2.1e-13;
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGGCTCGCGGTGAGCCCTGATCCGCTCGGGGTCCTCCAGTCCGCTGCTGCTGACG 62
Db 119 CAGGCTCGCGGTGAGCCATGATCCGCTCGGGGTCCTCCAGTCCGCTGCTGCTGACG 178
QY 63 CTGCTCATCCGCGGTCTTACGGTGTACGGGTCAGGCGCAGGATGCCGTAAGT 111
Db 179 CTGCTCATCCGCGGTCTTACGGTGTACGGGTCAGGCGCAGGATGCCGTAAGT 227

RESULT 8
LOCUS BQ715103 962 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8474655 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6310281
5', mRNA sequence.
ACCESSION BQ715103
VERSION BQ715103.1 GI:21854002
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 962)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Susan L. Sullivan, PhD.  
CDNA Library Preparation: ResGen, Invitrogen Corp  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM3732 row: f column: 10  
High quality sequence stop: 522.  
Location/Qualifiers  
1. .962  
/organism="Mus musculus"  
/db xref="taxon:10090"  
/clone="IMAGE:6310281"  
/clone\_lib="NIH\_MGC\_129"  
/lab\_host="NIH\_MGC\_129"  
/note="Organ: olfactory epithelium; Vector:  
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned  
unidirectionally. Primer: Oligo dT. Average insert size  
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this  
is a NIH\_MGC Library."  
is a NIH\_MGC Library." 139 t

SE COUNT 179 a 342 C 302 G 139 t  
IGIN

Query Match 10.9%; Score 102.6; DB 14; Length 962;  
Best Local Similarity 96.3%; Pred. No. 2.1e-13;  
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

3 CCGGCTCATCGCGGTGAGCCCTGATCCGCTCGGGCTCCCGAGTCCGCTGCTGCTGAGG 62  
|||||  
114 CAGGCTCATCGCGGTGAGCCATGATCCGCTCGGGCTCCCGAGTCCGCTGCTGCTGAGG 173  
|||||

63 CTGCTCATCGCGGTGAGCCCTGATCCGCTCGGGCTCCCGAGTCCGCTGCTGCTGAGT 111  
|||||

174 CTGCTCATCGCGGTGAGCCCTGATCCGCTCGGGCTCCCGAGTCCGCTGCTGCTGAGT 222  
|||||

SULT 9  
715834  
CUS  
FINITION

AGENCOURT\_8291643 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6308835  
5', mRNA sequence.  
BQ715834  
BQ715834.1 GI:21854733  
EST.  
YWORDS  
URCE  
ORGANISM

Mus musculus  
house mouse.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 965)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Susan L. Sullivan, PhD.  
CDNA Library Preparation: ResGen, Invitrogen Corp  
DNA Library Arrayed by: The I.M.A.G.E. Consortium  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM3728 row: j column: 04  
High quality sequence stop: 11  
High quality sequence start: 611.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 965)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Susan L. Sullivan, PhD.  
CDNA Library Preparation: ResGen, Invitrogen Corp  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM3735 row: d column: 14  
High quality sequence stop: 560.  
Location/Qualifiers  
1. .967  
/organism="Mus musculus"  
/db xref="taxon:10090"  
/clone="IMAGE:6311389"  
/clone\_lib="NIH\_MGC\_129"  
/lab\_host="NIH\_MGC\_129"  
/note="Organ: olfactory epithelium; Vector:  
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned  
unidirectionally. Primer: Oligo dT. Average insert size  
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this  
is a NIH\_MGC Library."  
is a NIH\_MGC Library." 146 t

BASE COUNT 165 a 305 C 350 G 146 t  
ORIGIN

Query Match 10.9%; Score 102.6; DB 14; Length 967;  
Best Local Similarity 96.3%; Pred. No. 2.1e-13;  
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

FEATURES  
source

Location/Qualifiers  
1. .965  
/organism="Mus musculus"  
/db xref="taxon:10090"  
/clone="IMAGE:6308835"  
/clone\_lib="NIH\_MGC\_129"  
/lab\_host="NIH\_MGC\_129"  
/note="Organ: olfactory epithelium; Vector:  
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned  
unidirectionally. Primer: Oligo dT. Average insert size  
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this  
is a NIH\_MGC Library."  
is a NIH\_MGC Library." 168 t

BASE COUNT 156 a 337 C 304 G 168 t  
ORIGIN

Query Match 10.9%; Score 102.6; DB 14; Length 965;  
Best Local Similarity 96.3%; Pred. No. 2.1e-13;  
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGGCTCATCGCGGTGAGCCCTGATCCGCTCGGGCTCCCGAGTCCGCTGCTGCTGAGG 62  
|||||  
Db 149 CAGGCTCATCGCGGTGAGCCATGATCCGCTCGGGCTCCCGAGTCCGCTGCTGCTGAGG 208  
|||||

QY 63 CTGCTCATCGCGGTGAGCCCTGATCCGCTCGGGCTCCCGAGTCCGCTGCTGCTGAGT 111  
|||||  
Db 209 CTGCTCATCGCGGTGAGCCCTGATCCGCTCGGGCTCCCGAGTCCGCTGCTGCTGAGT 257  
|||||

RESULT 10  
BQ950196  
LOCUS  
DEFINITION

BQ950196 967 bp mRNA linear EST 21-AUG-2002  
AGENCOURT\_8886884 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6311389  
5', mRNA sequence.  
BQ950196  
BQ950196.1 GI:22365674  
EST.  
KEYWORDS  
SOURCE  
ORGANISM

Mus musculus  
house mouse.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 967)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Susan L. Sullivan, PhD.  
CDNA Library Preparation: ResGen, Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM3735 row: d column: 14  
High quality sequence stop: 560.  
Location/Qualifiers  
1. .967  
/organism="Mus musculus"  
/db xref="taxon:10090"  
/clone="IMAGE:6311389"  
/clone\_lib="NIH\_MGC\_129"  
/lab\_host="NIH\_MGC\_129"  
/note="Organ: olfactory epithelium; Vector:  
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned  
unidirectionally. Primer: Oligo dT. Average insert size  
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this  
is a NIH\_MGC Library."  
is a NIH\_MGC Library." 146 t

BASE COUNT 165 a 305 C 350 G 146 t  
ORIGIN

Query Match 10.9%; Score 102.6; DB 14; Length 967;  
Best Local Similarity 96.3%; Pred. No. 2.1e-13;  
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

FEATURES  
source

Location/Qualifiers  
1. .965  
/organism="Mus musculus"  
/db xref="taxon:10090"  
/clone="IMAGE:6308835"  
/clone\_lib="NIH\_MGC\_129"  
/lab\_host="NIH\_MGC\_129"  
/note="Organ: olfactory epithelium; Vector:  
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned  
unidirectionally. Primer: Oligo dT. Average insert size  
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this  
is a NIH\_MGC Library."  
is a NIH\_MGC Library." 168 t

BASE COUNT 156 a 337 C 304 G 168 t  
ORIGIN

Query Match 10.9%; Score 102.6; DB 14; Length 965;  
Best Local Similarity 96.3%; Pred. No. 2.1e-13;  
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

y 3 CCGCCTCGCGGTGAGCCCTGATCCGCTCGGGCTCCCAAGTCGCTGCTGCTGACG 62
b 114 CAGGCTCGCGGTGAGCCATGATCCGCTCGGGCTCCCAAGTCGCTGCTGCTGACG 173
y 63 CTGCTCATCCCGCGGTCTACCGTGTACGGGTCAGGGCCAGGATGCCGTAAGT 111
b 174 CTGCTCATCCCGCGGTCTACCGTGTACGGGTCAGGGCCAGGATGCCGTAAGT 222

RESULT 11
OCUS BQ956014 976 bp mRNA linear EST 21-AUG-2002
EFINITION AGENCOURT_8800328 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6312433
CESSION BQ956014
ERISION BQ956014
EYWORDS 5', mRNA sequence.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 976)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13737 row: p column: 02
High quality sequence stop: 588.
Location/Qualifiers
1. .976
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned
unidirectionally. Primer: Oligo dT. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."
BASE COUNT 176 a 307 c 339 g 152 t 2 others
ORIGIN
Query Match 10.9%; Score 102.6; DB 14; Length 976;
Best Local Similarity 96.3%; Pred. No. 2.1e-13;
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

y 3 CCGCCTCGCGGTGAGCCCTGATCCGCTCGGGCTCCCAAGTCGCTGCTGCTGACG 62
b 139 CAGGCTCGCGGTGAGCCATGATCCGCTCGGGCTCCCAAGTCGCTGCTGCTGACG 198
y 63 CTGCTCATCCCGCGGTCTACCGTGTACGGGTCAGGGCCAGGATGCCGTAAGT 111
b 199 CTGCTCATCCCGCGGTCTACCGTGTACGGGTCAGGGCCAGGATGCCGTAAGT 247

RESULT 12
OCUS BQ714350 1003 bp mRNA linear EST 16-JUL-2002
EFINITION AGENCOURT_8480703 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6310055
CESSION BQ714350
ERISION BQ714350
EYWORDS 5', mRNA sequence.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1024)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

```

SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1003)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13731 row: 1 column: 24
High quality sequence stop: 506.
Location/Qualifiers
1. .1003
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned
unidirectionally. Primer: Oligo dT. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."
BASE COUNT 162 a 385 c 297 g 159 t
ORIGIN
Query Match 10.9%; Score 102.6; DB 14; Length 1003;
Best Local Similarity 96.3%; Pred. No. 2.1e-13;
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

y 3 CCGGCTCGCGGTGAGCCCTGATCCGCTCGGGCTCCCAAGTCGCTGCTGCTGACG 62
b 122 CAGGCTCGCGGTGAGCCATGATCCGCTCGGGCTCCCAAGTCGCTGCTGCTGACG 181
y 63 CTGCTCATCCCGCGGTCTACCGTGTACGGGTCAGGGCCAGGATGCCGTAAGT 111
b 182 CTGCTCATCCCGCGGTCTACCGTGTACGGGTCAGGGCCAGGATGCCGTAAGT 230

RESULT 13
BQ934161 1024 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8802223 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6336768
DEFINITION 5', mRNA sequence.
ACCESSION BQ934161
VERSION BQ934161.1 GI:22349544
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1024)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

Plate: LLAM13801 row: f column: 01  
High quality sequence stop: 485.  
Location/Qualifiers  
1. 1024  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6336768"  
/clone\_lib="NIH MGC 130"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: otcysts; Vector: pCMV-SPORT6.1.cdb;  
Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally.  
Primer: Oligo dT. Average insert size 1.95 kb.  
Constructed by ResGen, Invitrogen Corp. Note: this is a  
NIH MGC Library."  
SE COUNT 161 a 356 c 346 g 161 t  
IGIN

Query Match 10.9%; Score 102.6; DB 14; Length 1024;  
Best Local Similarity 96.3%; Pred. No. 2.2e-13;  
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

3 CCGGCTCGCGTGGAGCTGATCGCTCGGGGTCCCGAGTCGCTGCTGCTGACG 62  
147 CAGGCTCGCGTGGAGCTGATCGCTCGGGGTCCCGAGTCGCTGCTGCTGACG 206  
63 CTGCTCATCGCGCGTCTACGCTGTCAGGCGCAGGATGCCGTAAGT 111  
207 CTGCTCATCGCGCGTCTACGCTGTCAGGCGCAGGATGCCGTAAGT 255

RESULT 14  
LOCUS BQ714753 1051 bp mRNA linear EST 16-JUL-2002  
DEFINITION AGNCOUNT\_8480332 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6310056  
5', mRNA sequence.  
XCESSION BQ714753  
RSION BQ714753.1 GI:21853652  
WORDS EST.  
URCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Susan L. Sullivan, PhD.  
cDNA Library Preparation: ResGen, Invitrogen Corp.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
Cloning Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM13731 row: m column: 01  
High quality sequence stop: 541.  
Location/Qualifiers  
1. 1051  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6310056"  
/clone\_lib="NIH MGC 129"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: olfactory epithelium; Vector:  
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned  
unidirectionally. Primer: Oligo dT. Average insert size  
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this  
is a NIH MGC Library."  
SE COUNT 200 a 380 c 298 g 172 t 1 others  
IGIN

Query Match 10.9%; Score 102.6; DB 14; Length 1051;  
Best Local Similarity 94.8%; Pred. No. 2.1e-13;

Best Local Similarity 96.3%; Pred. No. 2.2e-13;  
Matches 105; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGGCTCGCGTGGAGCTGATCGCTCGGGGTCCCGAGTCGCTGCTGCTGACG 62  
DB 115 CAGGCTCGCGTGGAGCTGATCGCTCGGGGTCCCGAGTCGCTGCTGCTGACG 174  
QY 63 CTGCTCATCGCGCGTCTACGCTGTCAGGCGCAGGATGCCGTAAGT 111  
DB 175 CTGCTCATCGCGCGTCTACGCTGTCAGGCGCAGGATGCCGTAAGT 223

RESULT 15  
LOCUS W78398 477 bp mRNA linear EST 21-JUN-1996  
DEFINITION me78e08.r1 Soares mouse embryo NM013.5 14.5 Mus musculus cDNA  
clone IMAGE:401702 5' similar to gb:M65161\_rna2 Mouse pro-alpha  
(MOUSE);, mRNA sequence.  
ACCESSION W78398  
VERSION W78398.1 GI:1389042  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 477)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Gaisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:245470  
Seq primer: ETprimer  
High quality sequence stop: 340.  
Location/Qualifiers  
1. 477  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:401702"  
/clone\_lib="Soares mouse embryo NM013.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer 15'  
TGTTACCAATCTGAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTTTTTTTT  
T 3'), on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne  
State Univ., from 2 ); double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT73 vector. Library went through one round of  
normalization, and was constructed by Bento Soares and  
M.Fatima Bonaldo."  
BASE COUNT 86 a 148 c 149 g 94 t  
ORIGIN

Query Match 10.8%; Score 102.4; DB 14; Length 477;  
Best Local Similarity 94.8%; Pred. No. 2.1e-13;

